

Informationens skygge og foran

Informationskvalitet, informationsekspllosion og online kataloger

Poulsen, Claus

Publication date:
1994

Citation for published version (APA):

Poulsen, C. (1994). *Informationens skygge og foran: Informationskvalitet, informationsekspllosion og online kataloger*. Roskilde Universitet. Filosofi og videnskabsteori på Roskilde universitetscenter. 2.række, Samlinger Papirer om faglig formidling Nr. 36

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact rucforsk@kb.dk providing details, and we will remove access to the work immediately and investigate your claim.

Informationens skygge og foran

Claus Poulsen

Papirer om faglig formidling nr.36
Kommunikations Uddannelsen/RUC

Papirer om Faglig Formidling No. 36

Informationens skygge og foran
Informationskvalitet, informations-
eksplosion og online kataloger

Claus Poulsen

Institut for Datalogi, Kommunikation
og Uddannelsesforskning
Roskilde Universitetscenter
1994

Claus Poulsen: Informationens skygge og foran.
Informationskvalitet, informations-eksplosion og
online kataloger.
Institut for Datalogi, Kommunikation og Uddan-
nelsesforskning. Roskilde Universitetscenter.
1994.
(Papirer om Faglig Formidling No. 36).
© 1994 Forfatteren.
Omslag og layout: Forfatteren
Tryk: Roskilde Universitetscenters trykkeri og
Roskilde Fotosats & Trykservice
ISSN:0902-9001
ISBN:87-7349-264-7.

INDHOLDSFORTEGNELSE

Forord	i
Indledning	v
FØRSTE DEL:	
TEORETISKE OG EMPIRISKE FORUDSÆTNINGER	1
Kapitel 1	
Emnebeskrivelse og emnesøgning i bibliotekskataloger	3
1. Oversigt over emnesøgningens metoder	3
2. Bruges bibliotekskataloger til emnesøgning?	10
3. Forbedring af emnesøgningsmulighederne i bibliotekskataloger?	16
4. Brugernes ønsker til online kataloger	25
5. Sammenfatning og konklusioner	27
Kapitel 2	
Kontrollerede sprog versus naturlig tekst	29
1. Fordele og ulemper ved kontrolleret sprog og naturlig tekst	31
2. Indekseringskonsistens	35
3. Opsummering	46
Kapitel 3	
Ansætser til teoretisk analyse af emneindeksering	49
1. Hvad er et dokument emne?	50
2. Tolkninger, modeller og emner i klassisk modelteori	58
3. Informationsteoretiske betragtninger	74

INDHOLDSFORTEGNELSE

Kapitel 4	
Konsekvenser for emneindeksering i bibliotekskataloger	83
ANDEN DEL:	
PRAKTISKE ANVENDELSER	87
Kapitel 5	
Anvendelse af naturlig tekst til emnebeskrivelse i online kataloger	89
1. Analytiske kataloger	92
Kapitel 6	
Beskrivelse af paradoks-indeksering	105
1. Paradoks-kataloget på Danmarks Pædagogiske Bibliotek	107
Kapitel 7	
Evaluerings af et paradoks-indeksret katalog	115
1. Overflødigdigør PARADOKS-indeksering emneord i online kataloger?	121
2. Overflødigdigør PARADOKS-indeksering systematisk klassifikation i online kataloger?	123
3. Sammenligning af præciseringsmuligheder i et PARADOKS-indeksret og et traditionelt online katalog	125
4. Konklusion på evalueringen af PARADOKS-indeksering	132
Konklusion	137
English summary	143
Appendiks 1	
51 emneområder med encyklopædiske artikler	147
Appendiks 2	
Om statistikken	151
Appendiks 3	
Paradoks-indekserede poster	155

INDHOLDSFORTEGNELSE

Appendiks 4	
Emneord i paradoks-basen	159
Appendiks 5	
Systematiske klassifikationskoder i paradoks-basen	173
Appendiks 6	
Præciseringsmuligheder i et paradoks-indeksret og i et traditionelt katalog	179
Referencer	191
Stikordsregister	197

FORORD

PARADOKS-projektet er et forsøg på at skabe et alternativ til de traditionelle emnebeskrivelser i bibliotekskataloger.

Som ansat på et forskningsbibliotek er ét af mine mange gøremål, at foretage emnebeskrivelser af den anskaffede litteratur ved hjælp af klassifikations- og emneordssystemer.

Efterhånden begyndte jeg at tvivle på, at jeg selv ville kunne genfinde den værdifulde del af den litteratur, jeg selv indekserede. Faktisk fandt jeg, at den mest spændende del af litteraturen, nemlig den der specificerede sit videnskabsteoretiske udgangspunkt, slet ikke kunne beskrives med fornuft ved anvendelse af de traditionelle klassifikations- og emneordsværktøjer.

Resultatet blev, at litteratur, der udmærkede sig ved sin metode, emne, overblik eller videnskabsteoretiske angrebsvinkel, blev sat til side til jeg fik brug for det.

Arbejdsmetoden havde den ubetvivlelige fordel, at den gode litteratur reserveres til bibliotekets medarbejdere, idet den aldrig når at optræde i kataloget før den har mistet sin aktualitet.

Samtidig erfarede jeg, at kun en lille eksklusiv minoritet af brugerne overhovedet anvendte de tilføjede emneindekseringer.

Så jeg må vedstå, at jeg efterhånden fik udviklet det kætterske synspunkt, at ikke blot jeg, men også lånerne kunne have fordele af en anden fremgangsmåde til at beskrive litteraturens indhold end den traditionelle emneindeksering, der bygger på klassifikation og kontrollerede emneord.

Én af de vigtigste idéer fik jeg fra nogle ældre medarbejdere, der huskede store mængder titler på oversigtslitteratur og speciel litteratur, som de kunne hjælpe bibliotekets brugere og os andre med. Men hvorfor ikke gøre netop denne værdifulde type litteratur tilgængelig for brugerne via den nye teknologi, som i de år blev indført i bibliotekssektoren? Og hvorfor ikke gøre indholdsfortegnelser eller abstracts søgbare i kataloget i stedet for at foretage en på mange måder tvivlsom emneindeksering, ingen alligevel bruger?

For at afprøve denne idé blev der udviklet et mere eller mindre begrebsligt grundlag for, hvad der kom til at hedde PARADOKS-indeksering, hvor navnet er udsprunget af ønsket om at foretage PARAdigmespecifik DOKumentation:

Poulsen, Claus: Begrundelse for at anvende deltekstrepræsentation af metalitteratur til emnesøgning. Skriftserie fra Danmarks Pædagogiske Bibliotek, 6, 1987.

Poulsen, Claus: Subject Access to New Subjects, Specific Paradigms and Surveys: PARADOKS-registration. *LIBRI*, 40(3), 1990: 179-202.

For at kunne afprøve idéen skulle der foretages en registrering, således at det var muligt at sammenligne et PARADOKS-indekseret med et traditionelt katalog. Denne dobbelte registrering blev foretaget på Danmarks Pædagogiske Bibliotek i perioden 1987-89 og der blev foretaget en retrospektiv registrering for litteraturen i perioden 1985-87 i samarbejde med Bernard Cooper. Dele af det PARADOKS-indekserede katalog indgik i en litteraturlase om interaktive medier, distribueret i samarbejde med Dansk Interaktivt Mediecenter på kommerciel basis:

Poulsen, Claus: Paradox-registration af litteratur om emnet Interaktive Medier / Paradox-registration of literature on Interactive Media. A quarterly updated database on danish and english in the multimedia subscription Interaktive Medier, 1989-1991.

I løbet af dette samarbejde udvikledes et søgeprogrammel, der var specielt egnet til PARADOKS-indekseret litteraturregistrering.

Senere publiceredes et pilot-projekt til evaluering af PARADOKS-indeksering:

Poulsen, Claus: An indexing concept supporting subject access for innovation and creativity. In: Fugmann, Robert (ed.): Tools for Knowledge Organization and the Human Interface. Vol.2 Proceedings 1st International ISKO-Conference, Darmstadt, 14-17 August 1990. Frankfurt/Main, Indeks Verlag, 1991.

Poulsen, Claus: Overblik eller kaos? Om kvalitet og værdi i informationsformidlingen. *Information om Interaktive Medier*, 4, 1990: 18-20.

Nærværende bog er et forsøg på at samle PARADOKS-indekseringens forudsætninger og erfaringer, samtidig med, at det er et forsøg på at foretage en mere omfattende evaluering af denne indekseringsform.

Gamle kolleger fra Danmarks Pædagogiske Bibliotek, specielt Niels Skovgaard, Jan Rønnow, Bente Hvidberg, Mikael Søgaard Larsen og Bernard Cooper, har deltaget i inspirerende diskussioner og leveret kritik i løbet af PARADOKS-projektets udførelse.

Torsten Alf Jensen, Dansk Interaktivt Mediecenter har været en opmuntring til at arbejde med at udforme PARADOKS-indeksering til et færdigt produkt, og A/S Datacentralen har velvilligt stillet databaseprogrammet BRS-SEARCH gratis til rådighed for dette arbejde.

Irene Wormell, Danmarks Biblioteksskole, har udarbejdet en omfattende kritik og været en stor hjælp.

Karen Markey Drabenstott, University of Michigan har leveret gode idéer under en frokost-fremlæggelse.

Birger Hjørland fra Danmarks Biblioteksskole, Niels Erik Wille fra Statens Bibliotekstjeneste og Robin Cheesman fra Roskilde Universitetscenter leveret kritik, der har været frugtbar for projektet.

Og uden min nuværende kollega Pia Larsens store arbejde med at gennemlæse og korrigere det manuskript for skrivefejl, stavfejl og dårligt sprog, ville denne bog være vanskeligere tilgængelig, end det er nødvendigt.

Sidst - og ikke mindst - vil jeg udtrykke min glæde over gode og inspirerende diskussioner med kolleger og et godt arbejdsmiljø på min nye arbejdsplads, Roskilde Universitets Bibliotek.

INDLEDNING

Vi er blevet klar over, at vokser de rige landes produktion og forbrug, at vokser jordens befolkning, så skabes der store problemer i form af resurseknaphed og forurening. Materiel vækst er, fra at være et positivt begreb, blevet et begreb, der også forbindes med betydelige negative følgevirkninger.

Vor vækstafhængige kultur har til gengæld leveret et surrogat til den sørgeligt miskrediterede materielle vækst, nemlig viden.

Der synes ikke umiddelbart at være negative effekter af vækst i viden. Vækst i vor viden synes vi er godt.

Desværre er vækst i viden ikke direkte målelig, som vækst i den materielle produktion synes at være det.

Desværre ved vi også meget lidt om, hvordan man fremmer en sådan vækst. Hvordan sikrer man sig, at de midler der gives til produktion af viden, bliver anvendt til deres formål?

Det har været og er stadig politikernes og embedsmændenes problem, at viden ikke kan måles. Man er usikker på, hvordan man øger vidensproduktion, men de bevilgende myndigheder er - i praksis - tilsyneladende sikre på, hvordan vidensproduktion måles: Vidensproduktion måles i publikationer. Konsekvensen af denne styringsmekanisme er en publikationsekspllosion, hvor antallet af publikationer nu skønnes at blive fordoblet hvert fjerde år^a.

^a Hawkins 1994.

Sammen med tilsvarende udviklinger i andre sektorer i samfundet kaldes det også informationsekspllosionen.

Vækst i publikationer har ikke nødvendigvis noget med vækst i viden at gøre. Forskerne publicerer som aldrig før. Men lige som flere TV-kanaler med underholdning ikke nødvendigvis bidrager til, at folk får et morsommere liv, giver det ikke nødvendigvis mere viden, at der publiceres mere.

Informationsekspllosionen eller rettere publikationsekspllosionen har, som den materielle vækst, sine negative sider. Specielt giver det store problemer for formidlere af de videnskabelige publikationer. Informationsekspllosionen medfører nemlig, at det bliver mere og mere umuligt at følge med i originallitteraturen. I Chemical Abstracts gennemgås omkring 1,6 millioner dokumenter om året, og der udvælges mere end en halv million om året til abstracts og indeksering^a. Til Biological Abstracts bliver der udarbejdet emneindeksering og abstracts til godt 300.000 dokumenter om året.

Problemet er et problem for bibliotekerne.
Men er det et problem for forskerne?

Egentlig kan det være lige meget, om der publiceres 10.000 eller 100.000 artikler om året indenfor et fagområde. Den højt specialiserede forsker holder sig alligevel à jour med kolleger overalt i verden, får tilsendt fortryk eller artikler, tager på konferencer, og følger med i nogle tidsskrifter.
En sådan forsker har næsten ikke brug for biblioteker - og har aldrig haft det.
Eller har hun?

Også forskningsarbejdet har ændret sig.
Det bliver mere og mere almindeligt, at forskerne tvinges - via bevillingsstrukturen - til at ændre deres forskningsfelt jævnlige. Kun få kan fortsætte til pensionen med at arbejde videre med deres

^a Chemical Abstracts publicerer ikke statistik på antallet af gennemgåede dokumenter. Antallet er venligst oplyst fra Bob Gruebel, STN Help Desk. Antallet af indekserede artikler findes bl.a. angivet i CAS 1993.

speciale. Derved bevæger de sig væk fra deres vante informationskilder, bestående af bestemte tidsskrifter, kongresser og kontakter. Her har forskeren brug for biblioteket - eller kunne have brug for biblioteket.

Men også i situationen hvor forskeren ikke skal nyorientere sig, kunne bibliotekerne træde mere aktivt ind, nemlig som leverandører af "overraskende information".

Den snævert orienterede forsker står nemlig i den situation at kunne få alvorlige "fertiliteitsproblemer", da forskning også kræver nye og uventede informationer, der historisk set ofte kommer fra andre faglige miljøer.

Hvis bibliotekerne skal til at bidrage positivt til forskningsmiljøerne, bør dette have alvorlige konsekvenser for bibliotekernes praksis.

Hvis bibliotekerne ønsker at bidrage positivt til forskningsmiljøerne har det dybtgående konsekvenser for udvælgelse af litteratur til biblioteker.

Hvordan vælge, når ikke en gang den faglige specialist kan følge med i al litteraturen? Det gamle ideal om at opbygge en samling af den "vigtigste" eller "bedste" del af litteraturen er blevet uopnåeligt, da ingen ved, hvad der er det "vigtigste" eller det "bedste" - ingen har hele overblikket.

Løsningen *kunne* være, at bibliotekerne satsede på at anskaffe den litteratur, der søgte at skaffe sig et overblik over den store publikationsstrøm.

Hvis bibliotekerne ønsker at bidrage positivt til forskningsmiljøerne, har det også dybtgående konsekvenser for emnesøgningsfaciliteterne i bibliotekernes kataloger.

For ikke blot er den anskaffede litteratur præget af manglende overblik, med deraf følgende fald i resultatet af søgninger i bibliotekernes kataloger. Også litteraturens mangfoldighed vanskeliggør emnesøgninger i bibliotekernes online kataloger af acceptabel kvalitet, fordi bibliotekernes emneregistreringsmetoder slet ikke er indrettet på så store publikationsstrømme og hurtige faglige udviklinger, som vi er vidne til i dag. Resultatet er, at

brugerne læner sig mere og mere op ad de kommercielle informationstjenester.

Løsningen *kunne* være, at ændre emneregistreringen således at den præsenterede litteraturens indhold på en forståelig og anvendelig måde, så brugeren får en mulighed for at danne sig et billede af litteraturens indhold og kvalitet.

Hvis bibliotekerne skal have en funktion i fremtiden, hvor et væld af kommercielle informationstjenester præger markedet, bliver det ikke som forkølede eftersnakkere efter disse tjenester, men som et kritisk og overbliksgivende korrektiv til disse.

Dette kunne opnås ved at styrke anskaffelse og formidling af overbliksgivende litteratur og den marginale litteratur, der kan virke som et kvalitativt korrektiv til den dominerende litteraturs overvældende kvantitet.

PARADOKS-indeksering er et teoretisk og praktisk forsøg på at bidrage til løsningen af disse store problemer. Ideen er kort fortalt at lade emnesøgninger foregå i en del af onlinekataloget, nemlig i den del, der repræsenterer den overbliksgivende og den i emnemæssigt eller metodisk forstand marginale litteratur. Til gengæld skal denne litteratur så være fyldigt og forståeligt beskrevet gennem indholdsfortegnelser og andre dele af forfatterens egen emnebeskrivelse.

PARADOKS-projektet indeholder yderligere et forsøg på vise hvor god eller dårlig metoden er, ved at evaluere den foreliggende PARADOKS-indekserede bases emnesøgningsmuligheder.

FØRSTE DEL

TEORETISKE OG EMPIRISKE FORUDSÆTNINGER

EMNEBESKRIVELSE OG EMNESØGNING I BIBLIOTEKSKATALOGER

De videnskabelige bibliotekers vigtigste opgave er at tilbyde litteratur til brugeren, svarende til brugerens behov. Enhver undersøgelse af brugernes behov viser, at de først og fremmest ønsker at finde litteratur om et bestemt emne.

I det følgende skal vi se på de metoder, der anvendes til emnesøgning, for derefter at fokusere på den emnesøgningsstrategi, der består i at søge i bibliotekskataloger.

Vi skal se på undersøgelser af hvor meget - eller rettere hvor lidt - bibliotekernes kataloger anvendes til emnesøgning. Vi skal se på undersøgelser af hindringerne for at emnesøgninger kan foretages kvalificeret i bibliotekernes kataloger. Endelig skal vi se på hvilke ønsker brugerne selv har til online katalogerne.

1. OVERSIGT OVER EMNESØGNINGENS METODER

At finde litteratur om et emne kaldes emnesøgning.

Emnesøgning er næsten altid et kompleks af forskellige metoder eller strategier. I det konkrete tilfælde er disse strategier indvævet i hinanden.

Søgningen kan starte med opslag i en håndbog. Dette resultat påvirker brugerens emneopfattelse og søgningen gentages da med en ny strategi med den nye emneopfattelse - f. eks. henvendelse til en kollega eller genkaldelse af noget læst, hvorfra referencerne måske kunne bruges. Dette ændrer igen brugerens artikulation af sit emne

og en ny strategi anvendes - f.eks. orientering i et relevant tidsskrift, søgning i et online katalog eller der returneres til et nyt opslag i en håndbog eller lignende.

Hele denne proces er malende kaldt "berrypicking"^a. Lige som bjørnen får øje på ét bær, går hen og tager det, får øje på et nyt bær, går der hen og tager det - således er en emnesøgning karakteriseret af at være ikke-lineær. Hver enkelt strategi giver en ny retning på søgningen. Denne fremgangsmåde er meget langt fra forestillingen om den monostrategiske og lineære emnesøgning på online kataloger, hvor brugeren gennem stadig interaktion med online kataloget stædigt forfølger sit mål gennem tilretning af søgeprofilen.

Listen af strategier er lang, men her er et udpluk af de vigtigste, opstillet i en tilfældig orden:

- søgning i bibliotekskataloger
- opslag i løbende bibliografier over tidsskriftartikler, herunder citationssøgning
- "browsing", "græsning" eller "bladring"
- gennemgang af årgange af kernetidsskrifter
- opsporing af referencer i allerede læst litteratur (kædesøgning)
- søgning på forfattere, der har skrevet om et givet emne
- direkte henvendelse til emnespecialister
- personlige kontakter
- opslag i overbliksgivende litteratur

De enkelte elementer i denne liste bliver kommenteret i det følgende.

EMNESØGNING I BIBLIOTEKSKATALOGER

Hermed menes søgning på bestemte emner i den afgrænsede samling, et bibliotek udgør, via den fortegnelse eller det katalog, biblioteket har lavet over sin bestand af - hovedsagelig - bøger.

Søgningen kan foregå som emnesøgning ved at søge på emneklassifikation eller emneord. Søgningen kan også foregå som ordsøgning i titler m.v..

^a Bates 1989.

Denne strategi har den fordel, at litteraturen er umiddelbart tilgængelig i modsætning til de fleste andre søgestrategier, der henviser til såvel umiddelbart tilgængelig litteratur som litteratur, der skal skaffes fra andre biblioteker.

En anden fordel ved at foretage emnesøgninger i bibliotekskataloger er, at det er gratis for brugeren i modsætning til specielt online søgninger i løbende tidsskriftbibliografier, der ofte må betales af brugeren.

EMNESØGNING TIDSSKRIFTSBIBLIOGRAFIER

Emnesøgning i løbende bibliografier over tidsskriftartikler er meget anvendt.

Ofte har hvert fag én eller flere løbende tidsskriftbibliografier. Eksempler er Chemical Abstracts, Biological Abstracts, Zentralblatt der Mathematik und ihre Grenzgebiete. Andre tidsskriftbibliografier er bredere som Science Citation Index, og endelig findes der tidsskriftbibliografier uden en faglig, men med en sproglig, kulturel eller lokaliseringsmæssig afgrænsning, f.eks Dansk Artikelindeks og UNCOVER.

Normalt er disse bibliografier tildelt emneord, hvorved de ligner de sædvanlige bibliotekskataloger. Flere af disse tidsskriftbibliografier medtager endvidere den enkelte artikels abstract, hvilket er med til at gøre disse hjælpemidler både populære og anvendelige.

En helt anden opbygning repræsenteres af citationsindekserne, hvor emneafgrænsningen foregår implicit ved, at der søges på de dokumenter, der citerer et allerede fundet dokument.

Den i indledningen omtalte publikationsekspllosion giver ofte store problemer med at overskue, dels hvad den enkelte løbende bibliografi egentlig dækker, og dels hvilke bibliografier, der findes.

Hertil kommer, at den enkelte bibliografi ofte udkommer i flere forskellige udgaver med forskellige dækningsområder, tilpasset forskellige publikationsformer som trykte udgaver, diskette- eller CD-ROM-udgaver, suppleret med én eller flere online-udgaver placeret hos forskellige databaseværter^a. Dette har givet anledning

^a Poulsen 1993/Papir.

til en anden type bibliografier, bibliografier over løbende bibliografier, som dog hovedsagelig bruges af professionelle biblioteksarbejdere^a.

EMNESØGNING VED "BROWSING", GRÆSNING ELLER BLADRING

"Browsing" eller på dansk "græsning eller bladrning" er den mindre strukturerede eller mindre fokuserede søgning, der består i at man vurderer en samling litteratur, der tilfældigvis står i nærheden af det, man søgte. Med den elektroniske lagring af informationer om litteratur er der mange andre muligheder for begrebet "i emnemæssig nærhed af" end den endimensionale repræsentation af emneklassifikation på hylder - f.eks. litteratur, der indeholder samme emnespecifikke ord i titel, i abstract eller i indholdsfortegnelse.

Græsning foregår næsten i enhver emnesøgning - om ikke andet så ved enhver vurdering og sortering af søgeresultatet.

Betingelsen for at en sådan græsning kan foregå rimelig kvalificeret er, at der er noget at "græsse i" - at der er tilstrækkelig information til at ens vurderinger er noget værd. Dette er ikke noget problem for "græsning" på hylder, hvor hele dokumentet umiddelbart er til rådighed. Men det er i høj grad et problem i et traditionelt bibliotekskatalog, hvor brugeren blot får forevist en post med forfatter, titel, publikationsdata samt et par klassifikationskoder og emneord.

EMNESØGNING VED GENNEMGANG AF KERNETIDSSKRIFTER

Så snart de centrale tidsskrifter for et emne er indkredset, er en gennemgang af de relevante årgange en meget skattet metode. Metoden giver god relevans af det fundne materiale i forhold til brugerens interesser, da "græsningsvilkårene" er særdeles gode, fordi artiklen er fyldigt repræsenteret. Brugeren har virkelig mulighed for kun at vælge den relevante litteratur.

Problemet er at finde de relevante tidsskrifter samt at visse emner ikke har kernetidsskrifter i den forstand, at tidsskriftet hovedsagelig eller udelukkende beskæftiger sig med det ønskede emne.

^a Se f.eks. Gale 1992-.

EMNESØGNING VIA REFERENCER

Denne emnesøgningsteknik består i at gå tilbage i litteraturen - i modsætning til anvendelse af citationsindekser, hvor man går frem i litteraturen - ved hjælp af referencer.

Denne strategi er ikke afhængig af nogen form for formel emnebeskrivelse eller klassifikation.

Metoden er en meget populær emnesøgningsstrategi.

EMNESØGNING VIA FORFATTERSØGNING

Ofte modstilles forfatter- og emnesøgning. Men forfattersøgning er faktisk en effektiv og meget anvendt ingrediens i emnesøgning, når søgeren allerede kender et værk, og vil vide om værkets forfatter har skrevet mere om samme emne.

EMNESØGNING VIA DIREKTE HENVENDELSE TIL EMNESPECIALISTER

En sådan henvendelse foregår til en aktiv forsker indenfor søgerens interessefelt.

Metoden anvendes i vid udstrækning og er meget givende, fordi formidlingen af den enkelte forskers engagement til den, der søger litteratur om et emne, er en uvurderlig kvalitet, samtidig med at forskernes egne veje til litteratur er suppleret med personlige kontakter.

Metoden tilbydes også på de forskningsbiblioteker, hvor der ansat faglige specialister - fagreferenter - hvis opgave det blandt andet er at henvise til emnets faglige specialister.

EMNESØGNING VIA PERSONLIGE KONTAKTER

Forskerens medlemskab af faglige miljøer har været og er en uvurderlig kilde til forskerens litteraturforsyning gennem kollegers litteraturidéer, tilsendte preprints, deltagelse i konferencer etc..

Det er en meget anvendelig søgestrategi at kontakte "the old boys'" netværk.

Selvom de personlige kontakter er en meget vigtig ingrediens i mange emnesøgningsstrategier, er der mennesker, der af forskellige grunde ikke har disse kontakter. Enten fordi de ikke er medlemmer af et fagligt miljø, eller fordi de er begyndt at arbejde med nye emneområder, deres hidtidige faglige miljø ikke dækker.

EMNESØGNING VIA OVERSIGTSLITTERATUR

Litteraturoversigter eller metalitteratur er litteratur, der søger at give overblik over et emnes litteratur.

Overgangen mellem litteraturoversigter og almindelig litteratur er flydende når det drejer sig om videnskabelig litteratur, fordi de fleste værker også indeholder omtale af den litteratur, der er skrevet om emnet. Mere eller mindre indeholder litteraturen selv en emnesøgning foretaget af forfatteren.

Undertiden er der tale om helt brede dækningsområder som i Encyclopædia Britannica, bredt orienterende tidsskrifter som Science, Nature, New Scientist og Scientific American eller mere fagspecifikke leksika som Encyclopedia of Mathematics.

Der kan også være tale om specielle typer af litteratur, vi kalder dem specialbibliografier, artikler i håndbøger, "the state of the arts"- eller "review"-artikler.

Andre gange er der tale om samleværker, der for redaktøren repræsenterer et vigtigt sæt artikler om et emne enten i bogform eller i form af et temanummer i et tidsskrift.

Ved at søge efter emner gennem opslag i oversigtslitteratur og bibliografier genbruger man allerede foretagne emnesøgninger, oftest foretaget af en emnespecialist. Herved undgås meget unødvendigt arbejde og mange udgifter. Metoden er oven i købet, for områder der er dækket, undertiden den mest effektive enkeltstående søgemetode^a.

Metoden bliver yderligere aktualiseret af publikationsekspllosionen. Ved at indsnævre sin søgning til at gælde litteraturoversigter løses det problem, at en emnesøgning ofte giver alt for meget litteratur, på kvalificeret måde.

^a Ved konstruktion af en international bibliografi om informationssøgningens økonomi viste det sig, at ud af i alt ca. 7.000 forskellige referencer fandtes kun 176 fra online søgninger i tidsskriftbibliografier, mens hovedparten kom fra specialbibliografier. Davison et.al. 1988.

Nyere emneområder med betydelig interesse for forskning og offentlighed har ofte et så stort publikationsvolumen at selv bibliografier over originallitteraturen bliver vanskelige at håndtere^a. Således var der på to år publiceret flere tusinde artikler om risikoen for at elektriske felter skulle være sundhedsfarlige for mennesker. En publikation med tusinder af aktuelle artikler er - for de fleste - uinteressant. Ved at udvælge oversigtslitteraturen decimeres disse ofte flere tusinde artikler til omkring 100 aktuelle oversigter fra forskellige forfatteres perspektiver^b.

Derfor er forskerne i dag i højere grad tvunget til at anvende oversigtslitteratur som kilder til deres informationer. Selv den biografiske forskning er tvunget til at anvende oversigtslitteratur som kilder til deres forskning i nyere ekspansive emneområder, fordi det er uoverkommeligt at gennemlæse den originale litteratur^c.

RESUMÉ: SØGESTRATEGIER

Efter gennemgang af disse meget forskellige former for emnesøgning er det forhåbentlig klart, at emnesøgning foregår ved forskellige strategier, der kombineres i løbet af selve søgeprocessen, samt at der ikke er én strategi, der er den eneste anvendelige emnesøgningsstrategi.

^a Marcia Bates benytter søgning på metalitteratur i forbindelse med online søgninger og kalder søgninger, der anvender denne strategi for BIBBLE: "To BIBBLE (named after "bibliography") is to see if the searching one plans to do has already been done by someone else, especially in the preparation of a bibliography or literature search.... The essence of *BIBBLE* is to let your predecessor do the work. Many of the articles, reports or books covered in the online databases are themselves extensive bibliographies or state-of-the-art-reviews. In the sciences, a "review" often refers to a survey of a large body of current research literature, rather than to the evaluative reviews of a single book found in literary sources." Bates 1987.

^b Poulsen 1993/Højspænding.

^c Söderquist 1992.

2. BRUGES BIBLIOTEKSKATALOGER TIL EMNE-SØGNING?

Herefter er det et helt rimeligt at stille spørgsmålet:

Hvorfor overhovedet beskæftige sig med emnesøgning i<<<-
MÆRKE>>>

bibliotekskataloger?

Metoden udmærker sig jo ved ikke at være specielt god - måske tværtimod!

Et svar er, at emnesøgning i bibliotekskataloger - i modsætning til alle andre metoder - er bundet til det aktuelle biblioteks beholdning. Med andre ord kan brugeren få litteraturen hurtigt.

Et andet svar er, at metoden måske kunne forbedres ved at blive integreret i andre emnesøgningsstrategier - f.eks. ved at muliggøre søgning i oversigtslitteratur og håndbøger via kataloget eller ved at forbedre kataloget således, at det kunne bruges til kvalificeret "browsing" eller græsning via mere fyldige emnebeskrivelser af de enkelte dokumenter.

Mange studier af brugen af bibliotekskataloger er baseret på interviewundersøgelser og spørgeskemaer. Anvendelse af interviews og spørgeskemaer er nødvendige hvis man ønsker at afdække brugerens ønsker, grad af tilfredshed samt deres forslag til forbedringer. Nogle af disse studier synes at vise, at fremkomsten af adgang til online søgning ikke har haft nævneværdig indflydelse på de traditionelle metoders anvendelse i emnesøgningsprocessen^a.

Stribevis af sådanne studier viser også, at emnesøgning stadig er den hyppigste anvendelse af onlinekatalogerne^b, og Karen Markey opsummerede en række undersøgelser for Council for Library Resources (CLR) og fandt, at mellem 40% og 69% af søgninger i de medvirkende bibliotekers onlinekataloger var emnesøgninger^c.

^a Knightly 1979 og Bayer & Jahoda 1981.

^b Se Lewis 1987 for en oversigt.

^c Markey 1985.

Brugen af bibliotekskataloger kan også undersøges ved at analysere brugerens dialog med datamaskinen - såkaldte transaktionsdata. Her kan man undersøge, om brugeren foretager en emnesøgning ved at se, om søgningen foregår i felter, der indeholder emnedata. Disse metoder var også i brug i Karen Markey's undersøgelser for CLR og gav mellem 34% og 65% emnesøgninger^a, hvilket er konsistent med resultaterne af spørgeskema- og interviewundersøgelserne, refereret oven for.

Men den mest omfattende undersøgelse af transaktionsdata er Larson's undersøgelser af brugen af onlinekataloget i et stort amerikansk universitetsbibliotek, MELVYL^b.

UDVIKLINGEN AF BRUGEN AF MELVYL

Det var lidt af en overraskelse, da Larson ved hjælp af en storstilet analyse af transaktionsdata fandt ud af hvordan brugen af online kataloger havde udviklet sig over en længere periode^c.

Larson analyserede 15,3 millioner søgninger over en periode på knap 6 år i et stort amerikansk universitetsbiblioteks katalog - MELVYL.

Resultatet var, at andelen af emnesøgninger var signifikant faldende med gennemsnitlig 2,2% om året!

Lige så interessant er det, at sammenligne hvor meget kataloget bruges til søgninger på emner i helt bred forstand - "topical search" - i forhold til forfattersøgning, lokaliserings- og verifikationsopslag - "known item search"^d.

Larsons resultater er gengivet i Figur 1. Vi ser at forfattersøgning og verifikationsopslag optager en voksende del af søgningerne for

^a Markey 1985.

^b Det samlede katalog for universitetsbibliotekerne, der hører til University of California.

^c Larson 1991.

^d Emnelignende søgninger (topical search) defineres som søgning i samtlige emnebeskrivelsesfelter samt i det sædvanlige titelfelt. Verifikations- og forfatter-søgning (known items search) dækker over søgninger på samtlige forfatterfelter samt eksakte titler.

Larson 1991:figur 9.

FIGUR 1

Udviklingen af emne-, forfatter- og verifikationssøgninger på MELVYL gennem en 6-årig periode. Perioden jan84-apr85 er der ingen målinger, da systemet var under omlægning. (Larson 1991).

sidst i undersøgelsesperioden at være den dominerende anvendelsesform af kataloget. Bibliotekskataloget bliver altså i stigende omfang brugt til at finde ud af, om biblioteket har et bestemt dokument eller en bestemt forfatters værker, mens anvendelsen til emnesøgning aftager.

Bibliotekskataloget fungerer i stigende omfang blot som lagerkatalog.

Dette kan ikke tages til indtægt for, at brugernes interesse i emnesøgning er faldene, men at de - som adskillige brugerundersøgelser viser - i stigende omfang anvender andre emnesøgningsstrategier.

Men kunne dette fald i benyttelsen af kataloget til emnesøgning ikke tolkes som resultatet af, at flere og flere utrænede fik adgang til online kataloget efterhånden som antallet af brugerterminaler øgedes og netværksadgangen blev mere og mere populær? Kunne det ikke vise, at disse nye brugergrupper bare ikke ved, hvordan

man foretager emnesøgning? Problemet kunne så løses ved at forbedrede søgefaciliteter og ved at tilbyde mere undervisning i at anvende bibliotekskataloger.

Der er to forhold, der - iflg. Larson - imødegår denne tolkning af faldet i andelen af emnesøgninger.

For det første viser flere undersøgelser, at emnesøgning anvendes mindre, jo mere erfarne søgere der er tale om. Det gælder såvel biblioteksansatte som forskere og studerende. Det gælder såvel for kortkataloger som for online kataloger^a. Forklaringen på dette er, at erfarne søgere véd, at bibliotekskatalogernes emneregistrering er for ringe til at foretage egentlige emnesøgninger på. Erfarne søgere foretrækker søgninger i den ganske vist magre beskrivelse, der ligger i titler, undertitler samt forfattere.

For det andet viser den gennemsnitlige månedsvariation i brugen af emnesøgning, søgning på titelord og forfatter i Larsons undersøgelse, at ferieperioderne adskiller sig markant fra resten af året ved at have en lavere andel af emnesøgninger og højere andel af titelords- og forfattersøgninger end tilfældet re for resten af året - se Figur 2.

Hvis vi antager, at andelen af erfarne brugere (bibliotekspersonalet, forskerne og studerende langt henne i deres studieforløb) er større i feriemånederne juli-september og december-januar, foretager erfarne brugere mindre emnesøgning i bibliotekskataloger end de uerfarne - helt i overensstemmelse med de undersøgelser, der er nævnt ovenfor.

Larson tilføjer selv en alternativ forklaring gående ud på at hans resultater - såvel faldet i andelen af emnesøgninger igennem den 6-årige periode som det drastiske fald i ferieperioderne - kunne forklares med, at brugere med erfaring i emnesøgning anvender meget færre opslag i online kataloget end den uerfarne. Teoretisk set er denne tolkning også mulig, da der ikke eksisterer pålidelige

^a Markey 1980, Matthews og Lawrence 1984.

Larson 1991:Figur 8.

FIGUR 2

Gennemsnitlig månedsvariation på MELVYL for anvendelse af emnesøgning i forhold til søgninger på forfatter og titelord. Larson (1991).

undersøgelser, der rapporterer om *udviklingen* af brugernes problemer med at foretage emnesøgninger.

NORDISKE TEKNISKE HØJSKOLERS BIBLIOTEKER

Også på nordisk plan har der været foretaget undersøgelser af anvendelsen af online kataloger til emnesøgning. I løbet af 5 måneder i 1991/92 blev der foretaget en transaktionsanalyse af brugen af online katalogerne på de tekniske højskolars biblioteker i Norge, Finland og Danmark^a. Resultatet var, at emnesøgningerne, forstået som søgninger på UDK-tal, udgjorde mellem 2% og 10% - altså en forsvindende del - af det samlede antal søgninger.

SAMMENFATNING OG DISKUSSION

Den ekstremt lave anvendelse af emnesøgninger på de nordiske tekniske højskolars biblioteker - under 10% af søgningerne - kan

^a Høy Nielsen, Lomheim & Pasanen-Tuomainen 1992.

dels skyldes, at disse biblioteker hovedsagelig anvender decimal-klassifikation og næsten ikke emneord, og dels kan det skyldes, at de traditionelle tekniske og naturvidenskabelige fagområder har et standardiseret sprogbrug, så fritext søgning er mindre problematiske end i store dele af de humanistiske og samfundsvidenskabelige discipliner.

I Larsons store analyse af transaktions data på MELVYL finder han et markant fald i antallet af opslag på de emnebeskrivende felter, samtidig med at denne aktivitet er signifikant lavere i ferierne, hvor de yngre studerende ikke er til stede.

Faldet kan tolkes på to måder:

Brugerne undlader i stigende omfang at benytte online kataloger til emnesøgning.

Brugerne er blevet dygtigere til at foretage emnesøgning, således at de behøver færre opslag.

Den lavere aktivitet i ferierne kan tolkes ud fra samme præmisser:

De mere erfarne brugere anvender bibliotekskataloger sjældnere til emnesøgning.

De mere erfarne brugere behøver at anvende færre opslag til deres emnesøgninger.

Men "feriedykket" har også en tredje mulig forklaring, og det er, at bibliotekspersonalet kun foretager få emnesøgninger i ferieperioderne, fordi deres arbejde i disse mere stille perioder er koncentreret om de tekniske, interne processer på biblioteket - såsom magasinering og hylderevision.

For at finde den mest sandsynlige tolkning af disse iagttagelser, er det nødvendigt at se på, om der er andre forhold i anvendelsen af online kataloger til emnesøgning, der kunne sandsynliggøre én af tolkningerne.

3. FORBEDRING AF EMNESØGNINGSMULIGHEDERNE I BIBLIOTEKSKATALOGER?

Udskiftningen af forskningsbibliotekernes kortkataloger med online kataloger har været kolossalt resursekrævende. Penge, uddannelse og - ikke mindst - oceaner af tid er brugt til at medarbejderne har lært sig at håndtere den nye teknologi.

Det har været slående, at hele denne massive satsning af kapital og arbejdstid har været styret af ønsket om rationalisering og lettelse af interne processer i biblioteket. Selvfølgelig er der kommet online kataloger ud af det med de uomgængelige kvaliteter, at brugeren kan foretage søgninger på større litteraturmængder på kort tid, søge på alle ord i dokumentbeskrivelsen, kombinere søgekriterier, se om materialet er udlånt, få udskrifter af søgeresultater og søge fra andre lokaliteter end bibliotekets katalogsal. Alt sammen meget store fremskridt for brugeren af bibliotekets kataloger, men mere betinget af den forhåndenværende teknologi end af målrettet udvikling ud fra brugerens behov.

Stadig har brugerne vanskeligheder ved at anvende bibliotekets katalog til emnesøgning.

Dette ytrer sig ved at mellem 35% og 50% søgningerne ikke giver nogen fund først og fremmest fordi brugeren enten staver forkert eller ikke kan anvende katalogets emneklassifikation og emneord^a.

Men også det omvendte - alt for mange fund - er hyppigt et problem. Årsagen ligger i de store publikationsmængder. Efterhånden som bibliotekernes online kataloger vokser svarer flere og flere af emnebeskrivelser til brugernes søgeprofiler. Brugerne får et stort og ofte uoverkommeligt problem med at sortere de fundne poster - finde "den bedste litteratur"^b.

Normalt ønsker brugerne ikke at gennemgå hundreder af poster for at finde "den bedste litteratur". Larson fandt, at mens folk i

^a Bates 1977, Bates 1986, Markey 1983, Markey 1985, Larson 1991, Høy Nielsen, Lomheim & Pasanen-Tuomainen 1992.

^b Wiberley & Daugherty 1988.

gennemsnit af alle søgninger fik 77,5 poster, var gennemsnittet af fremviste poster 9.1 pr. søgning^a. Brugere kunne ikke på stående fod præcisere søgningen yderligere og orkede ikke at se det store antal poster igennem!

Mest interessant er det nok, at se på hvorledes éns søgning påvirker antallet af fund. Er der måder at søge på, der giver et mere overkommeligt antal poster end andre?

I Larsons undersøgelse af 15,3 millioner søgninger over en seksårig periode er antallet af fund for de forskellige feltspecifikke søgninger undersøgt og resultatet er gengivet i Figur 3, der viser, at det gennemsnitlige antal fund ved emneordssøgning er omkring 200, hvilket er uacceptabelt højt, mens titelsøgning i gennemsnit giver omkring 30 fund.

Larson 1991: Figur 11.

FIGUR 3

Det gennemsnitlige antal fund ved søgning på titel, forfatter og emneord på MELVYL gennem en 6-årig periode. Perioden jan.84-apr.85 er der ingen målinger, da systemet var under omlægning. Hentet fra (Larson 1991).

^a Larson 1986.

De forsøg, der er foretaget for at forbedre emnesøgning i bibliotekskataloger, kan - groft taget - opdeles i to kategorier: de, der satser på at forbedre selve søgeprocessen, og de, der søger at forbedre katalogernes indhold.

FORBEDRING AF SØGEPROCESSEN

Denne type forbedringer tæller områder som fonetisk stavning med den hensigt at undgå at søgerens stavefejl ødelægger emnesøgningen^a, semantiske associationsnetværk^b, vægtning af søgeord via brugerens udpegning af relevante dokumenter^c, muligheder for at søge på klassifikationsystemets verbale emnebeskrivelser og udnytte de hierarkiske strukturer i disse systemer^d, samt diverse anvendelser af ekspertsystemer eller "kunstig intelligens" til hjælp for selve søgeprocessen^e. På længere sigt er der vundet indsigt, der vil være en uvurderlig hjælp til at lette og kvalificere brugen af online kataloger, når der engang kommer kvalitet i deres emnebeskrivelser.

FORBEDRING AF EMNEBESKRIVELSEN

Den anden strategi til at forbedre emnesøgninger i online kataloger består i at forbedre selve katalogets indhold - i denne sammenhæng katalogets emnebeskrivelser.

Her har den dominerende fremgangsmåde bestået i, at forbedre online katalogernes emnebeskrivelser ved at udvikle klassifikations-systemer og emneordssystemer. En anden retning af forbedringer har været at benytte dele af dokumentets tekst - f.eks. indholdsfortegnelser - som emnebeskrivelser af dokumentet.

Som det fremgik af afsnittet om brugen af bibliotekskataloger (afsnit 1.2), har ingen af disse forbedringer haft gennemslagskraft nok til at have markant indflydelse på katalogernes anvendelse - eller måske har de indførte forbedringer ikke kunnet holde trit med

^a Walker 1987.

^b Bates 1986, Markey 1988.

^c Walker 1987.

^d Markey 1986/Class.

^e Salton 1989, Cooper 1988.

eller været irrelevante overfor de problemer, den voksende publikationsmængde påfører bibliotekernes online kataloger.

EMNESØGNINGERNES KVALITET

Men hvorfor er disse mange forsøg da ikke lykkedes?

Det mest oplagte svar er, at de hidtidige begreber og metoder til at analysere og forbedre emnesøgning i online kataloger er uanvendelige eller utilstrækkelige.

Ser vi på de metoder, der foreligger til vurderinger af emnesøgninger i bibliotekskataloger, holder de sig oftest til blot at undersøge sammenfald mellem søgerens og katalogets terminologi, eventuelt suppleret med en undersøgelse af om søgningen resulterer i et hjemlån som et kvalitetskriterium^a.

Men ovenstående metoder siger i sig selv ikke meget om kvaliteten af emnesøgningerne i bibliotekernes kataloger. Med de store publikationsmængder, vi har i dag, er brugerne ikke tilfredse med bare at få et mere eller mindre tilfældigt udpluk af emnets litteratur, der matcher med deres mere eller mindre præcist formulerede søgeprofiler. Det er ikke nok, at kræve at den fundne litteratur er relevant i forhold til emnerne - for der er for megen relevant litteratur. Brugerne er interesserede i den "bedste" litteratur om emnerne - hverken mere eller mindre.

Lancaster, Connel, Bishop og McCowan har foretaget en vurdering af emnesøgningers kvalitet i forbindelse med søgning i bibliotekskataloger^b.

Et af undersøgelsens formål er at finde sandsynligheden for at en erfaren søger finder den "bedste litteratur".

Et andet formål er at finde hvilke ændringer, der skulle til, for at øge denne sandsynlighed.

I undersøgelsen valgte man at definere "den bedste litteratur" som de referencer, der var anført i udvalgte artikler fra nyere encyklo-

^a Se f.eks. Bates 1977, Diodato 1986.

^b Lancaster, Connel, Bishop & McCowan 1991.

pædier. Valget er retfærdiggjort af, at artiklernes forfattere er emnespecialister, der antages at være udvalgt, fordi de har bidraget væsentligt til det pågældende forskningsområde.

På den måde skabtes 51 emneområder og 51 dertil svarende ekspertbibliografier, der antages at indeholde "den bedste litteratur" indenfor disse emner. Disse emner er bredt fordelt - lige fra mænds midtvejskrise til landkommuner i Rusland, præ-columbianske religioner og den matematiske disciplin, der kaldes kø-teori.^a

Undersøgelsernes formål er at finde ud af, om det er muligt med udgangspunkt i de anvendte artiklers tekst at foretage en emnesøgning i et stort bibliotekskatalog og genskabe artiklernes tilhørende bibliografier.

Bibliografierne blev rensat for tidsskriftartikler samt titler, der ikke fandtes i det pågældende online katalog. Herefter blev der foretaget egentlige emnesøgninger af to trænede søgere udelukkende på basis af teksten i de 51 encyklopædiske artikler *uden referencer*.

Resultatet blev, at den gennemsnitlige genfindingsprocent i de 51 emnesøgninger var 59% - vel at mærke beregnet ud fra, hvad det var muligt at finde i kataloget.

Umiddelbart et meget tilfredsstillende resultat. Men set i det perspektiv, at søgningerne var udført af meget trænede søgere, der havde erfaring med det anvendte emnedatasystem og havde forberedt sig godt på de konkrete søgninger, er det ikke så imponerende. Når dertil kommer, at en stor del af "den bedste litteratur" - altså litteratur der var omtalt i de encyklopædiske artikler - var tidsskriftartikler, vil resultatet være utilfredsstillende for den almindelige bruger.

Hvad enten man nu er tilfreds eller ej med at trænede professionelle søgere finder godt halvdelen af den tilgængelige "bedste litteratur", er det interessant at undersøge, om resultatet kunne være bedre.

Der blev foretaget to typer forbedringer.

^a Se Appendiks 1.

Først søgte man at forbedre emnesøgningernes kvalitet. Søgeprofilen blev udvidet med alle de muligheder, emnedatasystemet gav, selvom de lå i periferien af emnet.

Dernæst søgte man at forbedre emnebeskrivelsernes kvalitet ved at berige de eksisterende emnebeskrivelser med indholdsfortegnelser, abstracts og stikordsregistre fra den litteratur, der skulle findes, men endnu ikke var fundet.

FORBEDRINGER AF EMNESØGNINGERNES KVALITET

Den ene forbedring bestod i at udvide søgningen på de eksisterende bibliografiske posters emnebeskrivelser med alle de emneord, der kunne have en vis forbindelse med emnet. Derved øgedes antallet af *relevante* fund kun med 10%, mens antallet af fund i alt steg voldsomt.

Så ved at presse den eksisterende emnesøgningsteknik maksimalt kunne man kun forøge genfindingsprocenten af det relevante materiale til godt to tredjedele, mod til gengæld at få en masse materiale uden relevans, der skulle sorteres fra.

Konklusionen er, at der desværre kun kan gøres meget lidt for at forbedre emnesøgningen på grundlag af den eksisterende praksis for emnebeskrivelse. Semantiske associationsnetværk, fonetisk stavning, vægtede søgetermer, overskridelse af Booleske søgemetoder og såkaldt kunstig intelligens - ja alt hvad der kan tænkes indenfor sofistikerede interfaces vil højst hjælpe til at forøge genfindingsprocenten ganske marginalt for den trænede bruger. Grænsen i dette undersøgte tilfælde er 10%!

Én af årsagerne er selvfølgelig, at der er alt for få muligheder for at søge på emnebeskrivelser i de sædvanlige online katalogers poster. Normalt er der tre- fire klassifikationskoder eller emneord til at beskrive en bog på gennemsnitlig 300 sider, mens de meget anvendte tidsskriftdatabaser tilbyder søgning på et abstract på flere hundrede ord samt adskillige emneord for hver tidsskriftartikel af størrelsesordenen 10 sider. Dette giver selvfølgelig en meget dårlig genfinding i online kataloget samt oftest en meget ringe præcision, da der ikke er mulighed for at indsnævre søgeprofilerne mere end den sparsomme anvendelse af emneords- eller klassifikations-systemet tillader.

FORBEDRINGER AF EMNEBESKRIVELSERNES KVALITET

Den anden type forbedring bestod i, at benytte indholdsfortegnelser og emneindekser som beskrivelse af de tilsvarende dokumenters emner.

Undersøgelsen medregnede kun dokumenter, der ikke kunne genfindes ved hjælp af den eksisterende bibliografiske beskrivelse. Ved at tilføje dokumenternes stikordsregistre og indholdsfortegnelse opnåede man en forøgelse af genfindingsprocenten på godt 30%. Overraskende ville søgning i den fulde tekst kun bidrage til en yderligere forøgelse på knap 11%.

Forbedringer, bestående i at anvende stikordsregister eller indholdsfortegnelse, er altså langt mere virkningsfulde til at forøge genfindingsprocenten end de traditionelle forbedringer af søgesystem og interface.

Disse resultater understøttes af Karen Calhoun og Karen Markey's sammenligning af værdien af forskellige søgeelementer i online kataloger^a.

Studiet indeholder to undersøgelser.

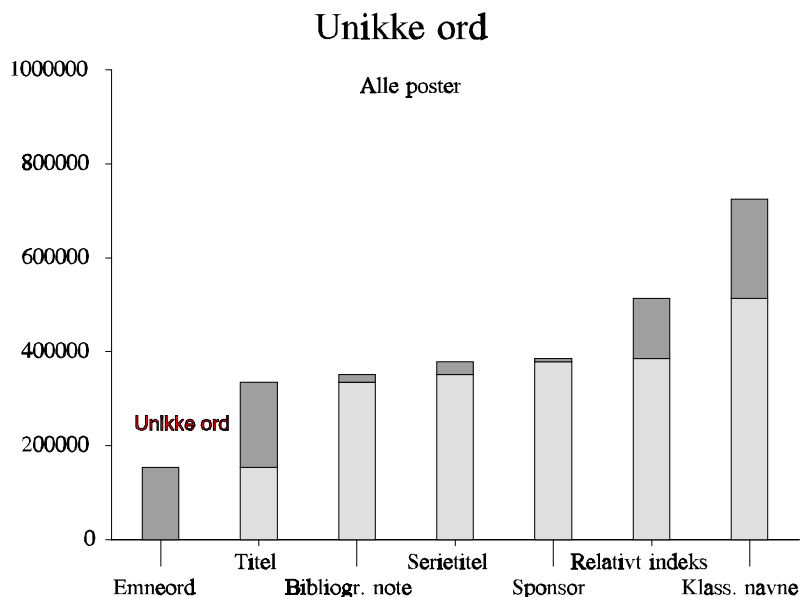
Den første består i analysere indholdet af 37.044 allerede eksisterende bibliografiske poster for emnebeskrivende ord. Formålet er at finde hvilke dele af den bibliografiske post, der bidrager mest til unikke emnebeskrivende ord. Som det er tilfældet i alle bibliotekskataloger forekommer emnebeskrivende ord i en hel række forskellige felter, nemlig emneordsfelter, titelfelter, felter med bibliografiske noter, seriefelter m.v..

Der blev foretaget en optælling af felternes successive bidrag med unikke ord.

Rækkefølgen valgtes at være: emneord, titel, bibliografiske noter, serier, sponsorer, relativt index fra det systematiske klassifikationssystem (DDC)^b, ord og noter fra klassifikationssystemet. Det betyder, at alle emneordenes forskellige ord er unikke, mens tit-

^a Markey & Calhoun 1987.

^b Dewey Decimal Classification system



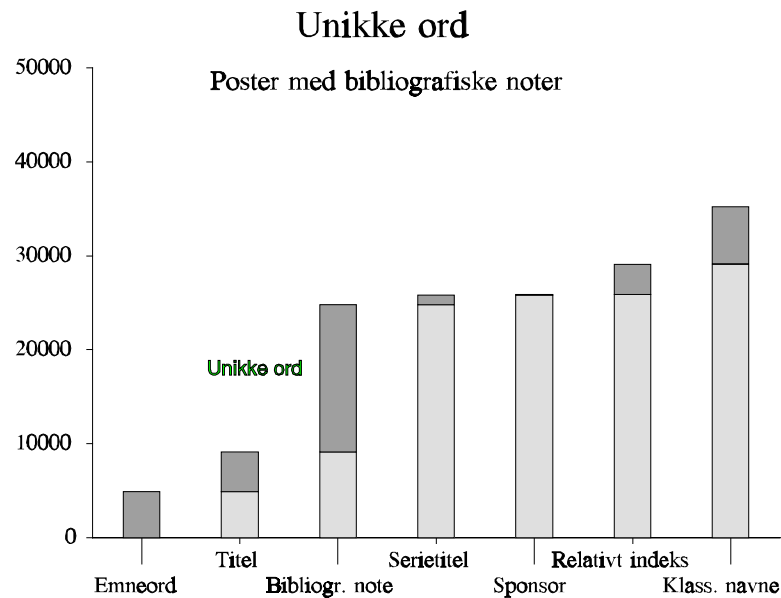
FIGUR 4

Successive bidrag af unikke ord fra emnebeskrivende felter i 37.044 poster. Skraverede felter er pågældende felts bidrag til unikke ord. Successionen begynder med emneord. Markey & Calhoun 1987:Tabel 2.

lernes ord kun er unikke, hvis de ikke findes blandt emneordene, ordene i de bibliografiske noter er unikke hvis de hverken forekommer blandt emneordene eller titelordene o.s.v..

Ikke overraskende fandt de, at emneord, titelord og klassifikationstermer fra DDC, dominerede bidragene ved denne successive fremgangsmåde - se Figur 4. Resultatet er netop ikke overraskende, fordi kun få poster overhovedet anvender felterne for bibliografiske noter, der dækker indholdsfortegnelser og abstracts, annotationer m.v..

Foretages analysen på de 1.010 ud af 37.044 poster, hvor felterne for bibliografiske noter anvendes, kommer resultatet til at se helt anderledes ud - se Figur 5. Her viser det sig, at langt den overvejende del af ordene til emnebeskrivelse kommer fra indholdsfortegnelser og abstracts, annotationer m.v..



FIGUR 5

Successive bidrag af unikke ord fra emnebeskrivende felter i de 1.010 poster med bibliografiske noter. Skraverede felter er pågældende felts bidrag. Successionen begynder med emneord. Markey & Calhoun 1987: Tabel 3.

RESUMÉ

De refererede undersøgelser viser, at brugerne har store vanskeligheder med at anvende online katalogerne til emnesøgning, idet de laver fejl i mellem en tredjedel og halvdelen af søgningerne. Samtidig har publikationsekspllosionen reduceret værdien af de søgninger, hvor det er lykkedes brugeren at anvende emneord, idet antallet af fund i store bibliotekskataloger ligger på et gennemsnit på ca. 200, hvilket er uacceptabelt højt.

Vi har også set, at søgninger på den "bedste del" af litteraturen kun kan forbedres væsentligt ved at berige de bibliografiske poster med indholdsfortegnelser, abstracts m.v. - ingen nok så avanceret udformning af søgegrænsefladen vil kunne gøre det.

Endelig viser en analyse af faktisk eksisterende bibliografiske poster, at det er fra indholdsfortegnelser, registre, abstracts og lignende tekster fra selve dokumentet, at der kan udtrages unikke emnebeskrivende ord - ikke fra emneordsfelterne.

NOGLE KONSEKVENSER

Selv om der ligger et ganske betydeligt potentiale til at få skabt større kvalitet i emnesøgninger i online kataloger ved at berige katalogets poster med indholdsfortegnelser, abstracts og anden naturlig tekst fra dokumenterne, er det desværre kun et potentiale. Det er ikke muligt ud fra disse resultater at konkludere, at bibliotekskatalogerne blot skal beriges med indholdsfortegnelser, indeks og anden tekst, taget fra dokumenterne.

I meget små samlinger og ved meget emnespecifikke ord er det helt uden tvivl en forbedring.

I online kataloger, der dækker store samlinger, kan det udvikle sig til et mareridt, hvor emnesøgning i stedet for at resultere i det nuværende uacceptabelt store antal fund på flere hundrede i gennemsnit vil resultere i flere tusinde fund.

4. BRUGERNES ØNSKER TIL ONLINE KATALOGER

I de foregående afsnit så vi på de metoder, der anvendes til emnesøgning, for derefter at fokusere på den emnesøgningsstrategi, der består i at søge i bibliotekskataloger. Vi så på hvor meget bibliotekskatalogerne blev anvendt til emnesøgning og hvilke hindringer, der lå for at hæve kvaliteten i emnesøgningerne.

De fleste af disse undersøgelser er foregået ved, at brugerens adfærd eller katalogets indhold er undersøgt. Målsætningen med at forbedre online katalogerne er ganske klar: at give brugeren et katalog, der er velegnet til emnesøgning. Det er derfor naturligt også at se på hvad brugernes ønsker og erfaringer er.

UNDERSØGELSER AF BRUGERNES ØNSKER

De til dato mest autoritative undersøgelser af brugeres behov blev foretaget i USA i begyndelsen af 1980'erne af en række organisationer med midler fra Council on Library Resources (CLR). Der er tale om undersøgelser af flere tusinde brugere og resultaterne er langt fra "opbrugt" til fornyelse af online kataloger.

I Online Computer Library Center's spørgeskemaundersøgelse blev brugere i 29 biblioteker bedt om at udvælge de fire vigtigste forbedringer af bibliotekernes online kataloger blandt en fortrykt liste af forslag. De hyppigst udvalgte ønsker blandt de mere end 20.000 svar var, at kunne se ord, der er forbundet med de aktuelle søgeord, og at få mulighed for at kunne se bøgeres indholdsfortegnelser, abstracts og stikordsregistre i løbet af søgeprocessen^a.

Behovet for at se ord, der er forbundet med éns søgeord, er en naturlig følge af, at emnebeskrivelsen i online katalogerne er baseret på et kontrolleret emneordssystem, som brugeren gerne vil kunne anvende så kvalificeret, som det er muligt. Dette er i dag praksis i de mange online kataloger og giver en vigtig sammenhæng i brugerens anvendelse af systemets emneordssystem, klassifikationssystem og søgning på titelord og anden fri tekst. I sin simpleste form er emneords- og klassifikationssystemet samt henvisninger gjort søgbare sammen med resten af katalogets indhold. Derved orienteres brugeren om emnedatasystemets muligheder såfremt det indeholder det anvendte søgeord.

Men brugerne ønskede også at kunne søge i indholdsfortegnelser, abstracts eller registre. Ved efterfølgende gruppeinterviews med brugerne blev disse ønsker udvidet til at bl.a. at dække titler på artikler i samleværker, forord, tekst på smudsomslag og anmeldelser.

Brugerne ønsker selvfølgelig dels at kunne finde litteraturen ved at blive "sluset ind" til katalogets kontrollerede emnebeskrivelser, og dels finde ud af, hvad bogen handler om.

PROBLEMER I RELATION TIL BRUGERNES ØNSKER

Det var gennemgående i undersøgelserne, at brugerne ønsker en voldsom berigelse af online katalogernes poster.

Det er allerede nævnt at sådanne berigelser ikke er særlig hensigtsmæssige i kataloger over store samlinger^b. De nuværende antal fund på i gennemsnit flere hundrede vil mangedobles. Eksempelvis

^a Markey 1983, Fayen 1983 og Markey 1985.

^b Se s.25.

fandt Byrne og Micco en forøgelse af genfindingen, ved at tilføje bl.a. indholdsfortegnelser til eksisterende bibliografiske poster, på mere end 300%!^a

Det hjælper i alle tilfælde ikke brugeren til at finde "den bedste litteratur".

Konklusionen er ikke at brugernes ønsker skal opgives, men at selve katalogets omfang og emnebeskrivelse skal tages op til samlet overvejelse.

5. SAMMENFATNING OG KONKLUSIONER

Vi har set, at emnesøgning er en højst sammensat og ulineær proces, bestående af mange forskellige søgestrategier, hvor emnesøgning i bibliotekernes kataloger er én strategi ud af mange - en strategi, der synes at blive brugt mindre jo mere erfarne brugerne er.

Der kunne være behov for, at nogle af disse strategier blev integreret, således at det blev nemmer at anvende de ellers separate strategier. For eksempel kunne søgning i oversigtslitteraturens indhold lægges ind i sammenhæng med søgning i bibliotekskataloger. Eller katalogernes information om hele eller dele af litteraturen kunne gøres mere fyldig, end der er tradition for nu, således at brugerne fik bedre vilkår til at browse i kataloget og vurdere dokumenternes relevans.

Vi har også set, at den voldsomme stigning i publikationsmængden vil øge de store vanskeligheder brugerne i forvejen har med at anvende online katalogerne til emnesøgning. Dermed er den mest sandsynlige tolkning af det prægnante fald i antallet af emneopslag i online katalogerne, at brugerne - efterhånden som de høster de dårlige erfaringer med bibliotekskataloger - prioriterer andre emnesøgningsstrategier højere.

^a Byrne og Micco 1988.

Forbedringer af bibliotekskatalogernes emnesøgningsfaciliteter kan foretages ved at ændre søgefaciliteterne eller ved at ændre katalogernes indhold.

Det er vist, at selvom udformningen af brugervenlige grænseflader, generaliserede vektor modeller og meget andet er vigtige, kan det ikke opveje det faktum, at det er de bibliografiske posters mangelfulde emnebeskrivelser, der er problemet.

Alle undersøgelserne peger på, at potentialet til at gøre bibliotekernes online kataloger til et værktøj for emnesøgning, ligger i at anvende indholdsfortegnelser, abstracts og lignende fri tekst som emnebeskrivelse.

Problemet er bare, at "den bedste" litteratur ofte drukner i den voldsomme publikationsmængde og de store fund.

Konklusionen er, at der ikke er nogen vej ud af problemet, hvis man fastholder det traditionelle krav, gående ud på, at hele online kataloget skal være til rådighed for emnesøgning.

Opgives dette krav er der derimod udveje.

Hvis emnesøgning kan foregå i en mindre del af bibliotekets katalog, begrænset til f.eks. encyklopædiske artikler, eventuelt suppleret med specialbibliografier og anden oversigtslitteratur, vil det kunne lade sig gøre at forbedre mulighederne for emnesøgning i kataloget ved at lade emnebeskrivelsen bestå i dokumenternes indholdsfortegnelse, stikordsregistre og lignende.

Men det betyder, at der skal søges på naturlig tekst - så kaldt "fritekst søgning".

Metoden løser nogle problemer, men der skabes også nye.

Er oversigtslitteraturen dækkende?

Ved søgning på naturlig tekst skal søgeren selv finde synonymer, endelser, vælge sprog og måske finde over- eller underordnede begreber. Kan man forvente det?

KONTROLLEREDE SPROG VERSUS NATURLIG TEKST

Emnebeskrivelse ved anvendelse af såvel kontrollerede sprog som naturlig tekst forekommer ikke blot i bibliotekskataloger, men indgår også i mange andre emnesøgningsstrategier.

Ved gennemgang af årgange af kernetidsskrifter søger man gennem titler og forfattere i naturlig tekst, og hvis tidsskrifterne indeholder emneregistre, vil de oftest være formuleret i naturligt tekst.

Søger man i tidsskriftbibliografier er der i regelen tale om, at der er anvendt kontrollerede emneord som søgeværktøj, mens titelord og abstracts er i naturlig tekst.

Endelig foregår der emneindeksering i oversigtslitteraturen i naturlig tekst på flere niveauer, idet såvel selve teksten som grupperingen af den refererede litteraturs emner i form af rubrikker eller afsnit er en emneindeksering.

De to sprog har fordele og ulemper. Disse fordele og ulemper er ikke generelle, men må ses i forbindelse med hver enkelt emnesøgningsmetode og tilhørende emnesøgningsværktøj.

Eksempelvis grupperes "undervisning af kvinder i naturvidenskab" sammen med "undervisning af minoriteter" i en stor oversigtsartikel^a.

Det betyder ikke så meget, når det læses i en publikation, hvor man efter anskuelse, kan tilslutte sig eller tage afstand fra forfatterens ideologi og videnskabssyn.

Hvis denne gruppering havde været anvendt i et online katalog, hvor det tilsvarende emneord ikke er synligt for brugeren, vækkes

^a Gabel, Kagan & Sherwood 1980.

brugerens kritiske sans ikke, og der er først muligheder for relevansvurderinger, når selve litteraturen skal gennemgås post for post.

I omkring 40 år har debatten bølget frem og tilbage mellem tilhængere af naturlige h.h.v. kontrollerede sprog til emneindeksering af dokumenter i bibliotekskataloger.

Og i omkring 40 år har disse indekseringssprog haft deres lidenskabelige tilhængere og modstandere. Så kraftigt er bølgerne gået, at nogle af de mere besindige forskere ligefrem må indlede deres rapporteringer på området med hvad den ene fløj og den anden fløj kan få ud af det pågældende arbejde^a og samtidig understrege, at resultaterne ikke kun kan tages til indtægt for den ene af fløjene. Og helt aktuelt har Robert Fugmann kritiseret anvendelsen af naturlig tekst til emneindeksering med nærværende PARADOKS-projekt som gennemgående eksempel. Kritikken tages op i det efterfølgende, men at emneindeksering med naturlig tekst er et følsomt emne, kan bedst illustreres med et lille citat fra hans konklusion:

Claims to the contrary which pretend to achieve genuine and equivalent or even superior *subject access* in a simple and cheap mechanical way, for example by merely selecting textwords and offering them for match with natural language query words in retrieval, are highly destructive because they jeopardize what exists and fail to provide a workable substitute. They merely provide *word access* rather than *subject access*. The ethical perspective of this problem is discussed by Wellisch.^b

Men før vi skal se på resultaterne af forsøg på at sammenligne de to indekseringssprog, vil det være nyttigt at fastslå, hvad der menes med de to begreber.

NATURLIG TEKST er i informationsvidenskaben en betegnelse for det sprog, der forekommer i de pågældende dokumenter. Ofte kaldes det også FRI TEKST i modsætning til kontrollerede sprog.

Naturlig tekst består af titel, abstract, uddrag fra teksten eller selve dokumentets fulde tekst, men kan også dække indholdsfortegnelsen,

^a Cleverdon et.al. 1966:1.

^b Fugmann 1991:126.

stikordsregistret eller andre knap så "naturlige" tekstdele af et dokument. Anvendelse af naturlig tekst til emneindeksering er blevet aktualiseret med anvendelsen af computere til kataloger og bibliografier, men stammer i sin moderne form fra begyndelsen af 50'rne.

KONTROLLEREDE SPROG er fundamentalt set en reduktion af de potentielt mulige sprog til emnebeskrivelser. Derved samles ord, der ellers adskilte sig af vidt forskellige årsager, rækkende fra forskelle i stavemåde, til synonymmer, til yderst komplekse forhold byggende på subjektive tolkninger eller til samling af flere underordnede begreber i ét overbegreb. Emnedatasystemer og klassifikationssystemer er eksempler på sådanne sprog.

Kontrollerede sprog har i forskellige udformninger altid været anvendt, og er stadig helt nødvendige i trykte emneindekser og i trykte bibliotekskataloger, der skal kunne anvendes til emnesøgning. Spørgsmålet er, om det er nødvendigt eller ønskeligt i et online katalog.

1. FORDELE OG ULEMPER VED KONTROLLERET SPROG OG NATURLIG TEKST

Kontrolleret sprog og naturlig tekst har begge deres fordele og ulemper, der skal gennemgås i det følgende.

EMNEINDEKSERINGENS LÆNGDE er formentlig den faktor i emneindekseringen, der forøger genfindning af brugbare dokumenter om et emne mest. Ved at tilføje ekstra søgemuligheder eller søgeord øges sandsynligheden for, at det enkelte dokument findes.

Da emneindeksering med naturlig tekst har betydelig flere muligheder for at blive omfangsrig end tilfældet er for de kontrollerede sprogs vedkommende, betyder det, at emneindeksering med naturlig tekst - alene på grund af emneindekseringens længde - tendentielt øger antallet af genfundne relevante dokumenter.

REDUNDANS er et ord for hvor meget "overflødig" information, der er i emnebeskrivelsen^a.

Et abstract er normalt meget mere specifikt end et sæt kontrollerede emneord og indeholder ofte flere ord for samme eller beslægtede begreber.

Redundans eller mangfoldighed øger sandsynligheden for at dokumentbeskrivelsen indeholder netop det ord, søgeren vil anvende. Samtidig er redundans vigtig for at kunne vurdere et dokument relevans ud fra en nuanceret beskrivelse.

Dette peger på, at emneindeksering med naturlig tekst over for emneindeksering med kontrollerede sprog har den fordel, at det øger antallet af genfundne relevante dokumenter og giver bedre muligheder for at udvælge de relevante dokumenter ud fra emnebeskrivelsen.

GENERELLE ELLER SYNTETISERENDE BEGREBER er almindelige i kontrollerede sprog - især i systematiske klassifikationssystemer, mens de er mere sporadisk forekommende i den naturlige tekst. Forskellige sprog og forskellige ord for det samme begreb samles under ét.

Og omvendt giver den hierarkiske opbygning i et klassifikations-system muligheder for at opsplitte og samle begreber efter behov. Denne styrke fremhæves ofte af tilhængere af kontrollerede sprog i angrebene på anvendelse af naturlig tekst til emneindeksering - for eksempel i Fugmann's kritik af PARADOKS-projektet^b.

Dette peger på, at emneindeksering med kontrollerede sprog øger antallet af genfundne relevante dokumenter, samtidig med at der tilbydes en forfinet søgemulighed i form af kontrollerede indsnævrings- og udvidelsesmuligheder af éns søgeprofil^c.

^a For en nærmere behandling af dette begreb henvises til side 76 ff.

^b Fugmann 1991:125.

^c Inden for nye emneområder mangler kontrollerede sprog dog ofte de syntetiserende begreber og visse overordnede begreber dannes også i naturligt sprog via ordenes former eller stammer. Så påstanden modificeres. Dette var nogle af resultaterne af en sammenlignende undersøgelse foretaget af Chartron et. al. 1989.

UTIDSSVARENDE MULIGHEDER FOR EMNEBESKRIVELSER forekommer i stort tal i kontrollerede sprog, hvis styrke netop er deres konservative begreber, der skal godkendes gennem beslutningsprocedurer, der koordinerer og involverer mange mennesker.

EKSEMPEL

Selv i Danmark, der i international sammenhæng bestemt ikke i 1970'erne og i begyndelsen af 1980'erne havde videnskabelig tradition for at betragte arbejdere, andre racer og indvandrere som handicappede, selv dér - på Danmarks Pædagogiske Bibliotek - anvendtes i 1980'erne emneordet "kulturelt handicapet" for litteratur, der beskæftigede sig med arbejderklassens børn, farvede og indvandrere^a

Da emneklassifikation kun sjældent foretages retrospektivt, betyder det, at nyskabende litteratur er vanskelig at finde ved hjælp af kontrolleret sprog.

Dette er næsten en modsigelse af emnebeskrivelsens formål i videnskabelige og demokratiske sammenhænge, hvor litteratur om gamle og velkendte emner oftest er nem at finde, mens litteratur om nye emner er vanskelig at finde.

Hvad skal man så foretage emneindeksering for, når de gamle emner godt kunne findes, fordi den i forvejen er velbeskrevet, mens de nye "ubeskrevne" emner stadig er ubeskrevne?

Derfor er den naturlige tekst de kontrollerede sprog overlegen til emnebeskrivelse og genfindning af litteratur, der er nyskabende.

MARGINAL LITTERATUR eller litteratur, hvis paradigme adskiller sig fra de dominerende forskningsretninger, er tit vanskelig at finde i de store publikationsmængder af - det siger sig selv - "normal litteratur". Årsagen til dette ligger - som det var tilfældet for den nye og banebrydende litteratur - netop i, at de kontrollerede sprog dannes i de dominerende forskningsparadigmers selvforståelse^b.

^a Titler som "Learning to labour : how working class kids get working class jobs" og "The education of poor and minority children : a world bibliography" begge fra 1981 er emneklassificeret under dette begreb (Danmarks Pædagogiske Biblioteks online katalog 940114).

^b Der henvises til Poulsen 1990:181-82 for en uddybning.

EKSEMPEL

Begrebet arbejderklasse fandtes ikke i 1970'erne i ERIC's thesaurus. Det bedste bud på begrebet var "cultural disadvantage". På det tidspunkt var arbejderklassen ikke et videnskabeligt begreb for den dominerende del af den amerikanske forskning. I firserne revideredes dette således, at arbejderklassen introduceredes i 1982(!) og "cultural disadvantage" blev lukket som emneord i 1980^a.

Omvendt indeholder naturlig tekst muligheder for dannelse af nye ord og ordsammensætninger til beskrivelse af de marginale begreber, metoder eller emner.

Da den marginale litteratur på mange måder er grundlag for den nyskabende litteratur, bør genfindning af den marginale litteratur prioriteres højt.

På dette område er naturlig tekst de kontrollerede sprog overlegne.

TVETYDIGE EMNEBESKRIVELSER forekommer først og fremmest i meget korte beskrivelser med naturlig tekst, hvor sammenhængen mangler. Dette har været den hyppigste kritik af anvendelsen af naturlig tekst til emnebeskrivelse - også i kritikken af PARADOKS-projektet. Fugmann kritiserer PARADOKS-projektet for at isolere ordene fra deres meningsgivende sammenhænge med det resultat, at hverken ordenes mening eller forfatterens mening med disse vil blive bevaret^b.

Selvom Fugmann glemmer, at der i PARADOKS-projektet er tale om hele indholdsfortegnelser og hele abstracts og ikke isolerede ord, har han da helt ret.

Kontrollerede sprog har en styrke ved - principielt - at være éntydige, defineret af vedkommende emnedatasystem.

Hvor emnedatasystemet er dækkende for dokumenternes emner, indeksørens forståelse samt brugerens forståelse - d.v.s. hvor vi kan tale om et klassisk centralperspektiv. Eksempler er "objektive" forhold som dokumenttypologi, visse geografiske og administrative emneaspekter samt klassiske begreber indenfor dominerende forsk-

^a Houston 1990.

^b Fugman 1991:124.

ningsparadigmer - i disse tilfælde er de kontrollerede sprogs emnebeskrivelser éntydige^a.

Hvad Fugmann og andre kritikere af anvendelse af naturlig tekst til emneindeksering glemmer, er alle de andre tilfælde, hvor emnerne indgår i *forskellige* og konkurrerende paradigmer eller perspektiver - en situation, der er konstituerende for moderne erkendelse. I sådanne tilfælde har emnerne ikke en indlysende "objektiv" karakter. Og hvad de også glemmer er, at selvom den naturlige teksts emnebeskrivelser i *søgeprocessen* ofte er flertydige, så ophæves disse flertydigheder sædvanligvis ved læsning af de fundne posters emnebeskrivelserne i form af abstracts eller indholdsfortegnelser.

RESULTAT

Der kan ikke konkluderes entydigt ud fra ovenstående betragtninger. Det samme er tilfældet for empiriske undersøgelser af de to emneindekseringsformer. Til trods for en ganske stor eksperimentel aktivitet på området, er det ikke lykkedes at pege på om kontrollerede sprog eller naturlig tekst er den bedste metode til emneindeksering i bibliotekernes online kataloger.

Hidtidige resultater peger på at de to sprog supplerer hinanden^b.

2. INDEKSERINGSKONSISTENS

Det er en udbredt antagelse, at litteratursøgningers resultat er afhængig af at emneindekseringen er konsistent, d.v.s. at to indeksører, der stilles over for samme dokument foretager samme indeksering af dette dokument's emne. Umiddelbart forekommer det også logisk at antage, at sikring af ensartethed i emnebeskrivelsen er én af forudsætningerne for at dokumenterne kan genfindes.

^a Denne problematik er - på forskellig vis - et af hovedemnerne for det efterfølgende afsnit.

^b Se Henzler 1978 og Lancaster 1989, der indeholder en kritisk litteraturgenmængang.

Det er forventeligt, at de faktorer, der påvirker indekseringskonsistens er mangfoldige og komplekse.

Allerede i 1969 opregnede Zunde og Dexter 25 forskellige faktorer, der påvirkede indekseringskonsistens^a. Spekeret gik lige fra forfatterens sprog, over emnets karakter og indekseringssystemet til lokal-ets luftfugtighed og temperatur. Senere er flere kommet til - for eksempel organisatoriske forholds betydning for indekseringsprocessen.

Der er flere arbejder, der giver en oversigt over emnets mangfoldighed. Disse bliver intensivt benyttet i den efterfølgende oversigt^b.

Af hensyn til forståelsen af en så sammensat problematik, er det hensigtsmæssigt med en gruppering af de indgående parametre. Denne grovopdeling består i:

VARIATIONER AF INDEKSERINGSOBJEKTETS PARAMETRE, parametre, der stammer fra indekseringsobjektet - dokumentet eller den anvendte dokumenterstatning (d.v.s. titel, abstract eller hele teksten)

VARIATIONER AF INDEKSERINGSVÆRKTØJETS PARAMETRE, parametre, der kendetegner indekseringssproget, herunder indekseringsregler

VARIATIONER AF FAKTORER, DER PÅVIRKER INDEKSØREN, parametre, der beskriver indeksøren, herunder uddannelse og miljø.

Vi forestiller os en situation, hvor dokumentet indekseres alene på grundlag af den velkendte titel "Fysik". Indekseringsreglerne er således udformet, at hvert værk skal tilordnes én og kun én beskrivelse af dokumentets hovedemne. Emnedatasystemet har kun ganske få ingredienser, hvoraf én er "fysik", og disciplinen "fysik" er velafgrænset og velfunderet. I denne situation vil der med stor sandsynlighed være tale om, at den resulterende emneindeksering

^a Zunde & Dexter 1969.

^b Leonard 1977 og Markey 1984.

er konsistent. Men til gengæld er denne emneindeksering ganske uden informationsindhold - den overrasker ikke, og kan ej heller betegnes som en berigelse.

Med en så "stabil" situation som udgangspunkt vil vi nu se på isolerede påvirkninger af indekseringskonsistensen, stammende fra variationer af de tre grupper parametre^a.

VARIATIONER AF INDEKSERINGSOBJEKTETS PARAMETRE

Fastholdes alt andet end de parametre, der bestemmer indekseringens objekt, dokumentet eller dokumenterstatningen, forekommer det rimeligt at antage, at jo længere dokumentet eller dokumenterstatningen er, jo mere er der at være uenige om. Det er derfor ikke overraskende at finde, at indekseringskonsistensen falder signifikant hvis indeksøren indekserer på grundlag af den fulde tekst i stedet for at anvende dokumentets abstract og titel^b eller at indekseringskonsistensen er faldende med stigende længde af det indekserede dokument^c.

Derimod er det overraskende, at dokumentets læselighed ikke synes at have indflydelse på indekseringskonsistensen^d.

Emnets art og specificitet må influere på indekseringskonsistensen. Men denne påvirkning har kun mening set i samspil med såvel indeksørens egen viden og indekseringssystemets egnethed til beskrivelse af pågældende emneområde. Der er påvist en svag

^a En sådan metode vil ikke nødvendigvis kunne beskrive den virkelige situation, fordi de forskellige parametre påvirker hinanden med det resultat, at de enten forstærker eller formindsker deres indflydelse på indekseringskonsistensen. Og endelig ligger der bag en sådan metode den antagelse, at den virkelige situation er i ligevægt og opfører sig "ikke-kaotisk" i den forstand at parameter-værdier, der ligger meget tæt op ad hinanden resulterer i indekseringskonsistenser, der ligger tæt op ad hinanden.

^b Tell 1969.

^c Rodgers 1969, Fried & Prevel 1966 og Leonard 1975.

^d Zunde & Dexter 1969. Læseligheden er her ikke målt som forståelsesvanskeligheder, men udelukkende ud fra ordlængder og sætningslængder ligesom det danske LIX-tal.

sammenhæng mellem indeksørernes emnespecifikke viden og indekseringskonsistensen i to af undersøgelserne, mens der er påvist manglende sammenhæng i tre andre undersøgelser, hvor indeksører uden emnespecifik viden indekserede lige så konsistent^a.

Det er vanskeligt at afgøre, hvad der egentlig måles i sådanne undersøgelser. Dokumenterne kendes ikke. Såvel indeksørernes emnespecifikke viden som indekseringssystemets egnethed er vanskeligt at beskrive - og de er heller ikke tilstrækkeligt beskrevet. Der er simpelthen for mange parametre, der ikke kan holdes konstante i sådanne målinger. Derfor er resultatet også forventeligt uklart.

VARIATIONER I INDEKSERINGSVÆRKTØJETS PARAMETRE

Indekseringsværktøjet består af emnedatasystemet og de dertil hørende indekseringsregler og arbejdsforskrifter.

Indekseringssystemet kan være et kontrolleret eller et ukontrolleret system, d.v.s. enten skal indeksøren vælge blandt på forhånd fastlagte emnebeskrivelser eller indeksøren kan frit vælge enten fra teksten eller anvende eget sprog. Det må forventes, at det i almindelighed gælder, at anvendelse af kontrollerede emnedatasystemer giver større indekseringskonsistens end det er tilfældet ved anvendelse af ukontrollerede emnedatasystemer. Dette er i hovedsagen også bekræftet i de foretagne undersøgelser^b, mens en enkelt publikation dog når til et andet resultat^c.

Det må endvidere forventes, at en forøgelse af antallet af mulige valg af indekstermer ved et mere specifikt indekseringssprog, vil bevirke en lavere indekseringskonsistens, hvilket er bekræftet empirisk^d.

Stigende indekseringsdybde kunne man umiddelbart forvente ville betyde faldende konsistens, fordi jo flere indekstermer, der skal vælges til at beskrive et dokument, jo flere muligheder for inkonsi-

^a Disse undersøgelser resumeres i Markey 1984:164.

^b Tarr & Borko 1974, Herr 1970.

^c Slamecka & Jacoby 1965.

^d Leonard 1975:150, og 152, Reich & Biever 1991.

stens. Hvis der kun må vælges én indekstern til at beskrive et dokument, må vi forvente højere konsistens, end hvis der mulighed for fem. Men hvis der ingen begrænsninger er, kan den enkelte indeksør nå igennem alle emnerne, og der må forventes en større konsistens igen, simpelthen fordi rækkefølgen af valgene ikke mere spiller nogen rolle. Derfor må der forventes modstridende resultater af undersøgelserne, hvilket også er tilfældet^a.

VARIATIONER AF FAKTORER, DER BESKRIVER INDEKSØREN

Denne gruppe parametre til bestemmelse af indekseringskonsistens, er den mest uhåndgribelige, talrige, dårligst undersøgte og - formentlig - også den væsentligste.

Ud af Zunde Dexter's oprindelige liste på 25 faktorer, der påvirker indekseringskonsistensen, hører de 16 til denne gruppe^b. Der er foretaget empiriske undersøgelser af konsistensens afhængighed af 2 af disse. Senere har Livonen foretaget en undersøgelse af indekseringskonsistensen ved organisering i lokale klassifikationscentre i stedet for organisering i et centralt klassifikationscenter^c.

I alt er der altså kun foretaget undersøgelse af 3 indeksørafhængige faktorerets betydning for emneindekseringskonsistens. Det drejer sig om indeksørens vidensniveau og rutine, om indeklimaet i indekseringslokalet samt indekseringens organisation.

Umiddelbart må det forventes, at indeksørens vidensniveau og rutine påvirker indekseringskonsistensen positivt. Men dels har vi allerede - i forbindelse med variationer i indekseringsobjektets parametre - nævnt, at der ikke har kunnet påvises nogen signifikant positiv korrelation mellem emnespecifik viden og indekseringskonsistens. Dels har emnespecialister en tendens til at se al litteratur ud fra netop deres speciale^d. Og dels synes der alt i alt

^a Reich & Biever 1991, Pötzch 1980 og Leonard 1975, cit. efter Leonard 1977:32.

^b Zunde & Dexter 1969.

^c Livonen 1990.

^d MacMillan & Welt 1961.

ikke at være empirisk entydigt belæg for at erfarne indekserer indekserer mere konsistent end uerfarne indekserer^a.

Det fysiske arbejdsmiljøs indflydelse er blevet undersøgt af Zunde og Dexter^b. Resultatet blev, at det ikke var muligt at påvise sammenhæng mellem variationer i lokalets temperatur og indekseringskonsistensen. Denne undersøgelse er den eneste publicerede, der beskæftiger sig med det fysiske arbejdsmiljø. Der mangler altså meget før det fysiske arbejdsmiljø er beskrevet - specielt støj, lys- og strålingspåvirkning, ergonomi.

Betydningen af indekseringens organisation, hvor det udover arbejdets institutionelle organisering, handler om afbrydelser, informationsfaciliteter, tidspres, herunder fragmentering af arbejdsdagen eller sammenhæng mellem arbejdets forskellige dele, er en anden vigtig faktor til bestemmelse af indekseringen af dokumenter. Den eneste undersøgelse, der foreligger indenfor dette område er Iivonens undersøgelse af 10 forskellige forskningsbibliotekers indeksering af samme dokumenter med samme kontrollerede emnedatasystem (UDC)^c. Hun fandt bl.a., at indekserens opfattelse af sine (lokale) brugeres behov påvirkede indekseringen i retning af at tilfredsstille disse brugere! Men det sænkede altså også konsistensen mellem de forskellige organisationers indeksering, hvilket er af betydning for nationale bibliotekskataloger, der sammenstykket af de enkelte bibliotekers kataloger^d.

Tilbage er der et væld af u-undersøgte faktorer som indekserens alder, modenhed, personlighed, personlige historie, egnethed til indekseringsarbejdet, humør og sundhedstilstand, psykologiske tilstand, interesser, metavidenskabelige viden og standpunkt og meget andet.

^a Markey 1984:161, 164 og 166.

^b Zunde og Dexter 1969.

^c Universal Decimal Classification system.

^d Iivonen 1990).

HVAD PÅVIRKER INDEKSERINGENS KONSISTENS?: ET RESUMÉ

Skal der foretages et resumé af det foregående kan vi med en vis grad af sikkerhed påstå, at hvad angår selve indekseringens objekt - altså dokumentet eller dokumenterstatningen, er det en fordel jo kortere dokumentet eller dokumenterstatningen er. Titel eller abstract synes at give optimal konsistens.

Hverken dokumentets emne eller dets læselighed er påvist at have indflydelse på konsistensen af emneindekseringen.

Forøgelse af indekseringskonsistens, stammende fra valg af indekseringsværktøjer og -regler, skyldes valg af kontrollerede i stedet for ukontrollerede emnedatasystemer samt valg af systemer med brede i stedet for detaljerede emneangivelser.

Endelig er det vist, at indeksering der foregår på lokale indekseringscentre sænker indekseringskonsistensen, mens der i øvrigt ikke er foretaget undersøgelser af indeksørens betydning for indekseringen - bortset fra, at indekseringslokalets temperatur ikke synes at have indflydelse på indekseringens konsistens.

Det kan konkluderes, at bortset fra ganske få aspekter, har vi ingen viden om hvordan indekseringskonsistensen optimeres. Specielt har vi ingen viden om indeksørens indflydelse.

Herudover skal det pointeres, at hele den form, der anvendes til at analysere et så komplekst fænomen som indekseringskonsistens, formentlig er inadekvat, fordi den samlede påvirkning af emneindekseringen er andet og mere end summen af enkeltpåvirkningerne.

Endelig viser det ubestemte antal faktorer, at der mangler en egentlig teoretisk begribelse af fænomenet, som kunne vejlede til opstilling af modeller for empirisk afprøvning.

MÅL FOR INDEKSERINGSKONSISTENS

Vi har hidtil set på nogle af de faktorer, der påvirker emneindekseringens konsistens. Selvom vi ikke ved så meget om, hvordan man kan optimere denne konsistens, kunne det være, at det heller ikke gjorde noget. Selvfølgelig ville det skabe problemer at det kontrollerede éntydige sprog anvendes flertydigt. Men hvor meget gør det egentlig?

Det vil vi se på i det efterfølgende.

Det første spørgsmål, der melder sig er, hvad er vi tilfredse med: 50%, 75%, 90% eller 95% konsistens?

Svaret på dette spørgsmål afhænger selvfølgelig af, hvordan konsistensen måles.

Overraskende nok anvendes der meget forskellige mål for konsistens i den litteratur, der er refereret i det foregående.

Det gør ikke så meget, hvis man er interesseret i at måle konsistensens variation som funktion af forskellige variable. Der er da tale om et relativt mål indenfor den enkelte målerække. Det betyder til gengæld, at de forskellige undersøgelsers mål for konsistensens størrelse vanskeligt kan sammenlignes.

Grundlæggende har konsistensmål dog samme funktion: de skal være et mål for hvor meget af indekseringen, der er fælles for de betragtede repræsentationer af samme dokument.

Er der tale om sammenligning af to dokumenter anvendes sædvanligvis det såkaldte Tanimoto-, Spencer- eller Hooper-mål (selv navngivningen er man uenig i)^a, der defineres som forholdet mellem antallet af fælles emneindekseringer og antallet af indekseringer i alt.

Hvis

a= antallet af emneindekseringer anvendt af den første indeksør

b= antallet af emneindekseringer anvendt af den anden indeksør og

c= antallet af emneindekseringer anvendt af begge indeksører

fås ovennævnte konsistensmål:

$$K_1 = \frac{c}{a+b-c}$$

I litteraturen har konsistensmålet været meget anvendt.

^a Tanimoto mål er anført i Henzler 1978. Det tillægges Spencer i Rolling 1981, men Hooper-målet nu synes at være standardnavnet (se f.eks. Leonard 1977 og Iivonen 1991).

$$K_2 = \frac{2c}{a+b}$$

For $a=12$, $b=10$ og $c=8$ giver de to mål henholdsvis $K_1=8/14= 57\%$ og $K_2=16/22= 73\%$.

Det giver selvsagt vanskeligheder, hvis man ønsker at sammenligne de grader af konsistens, der måles af forskellige forfattere.^a

Når dertil lægges, at kun få af forfatterne redegør for hvorledes de optæller éns deskriptorer - f.eks. om det er tale om identiske afbildninger af hele ordstrengen eller ord for ord, om synonymer tillades og om tages der hensyn til variationer af endelser m.v. - så er det yderligere vanskeliggjort at sammenligne resultaterne af de foretagne målinger af indekseringskonsistens^b.

Resultaterne af de foretagne målinger af indekseringskonsistens viser, at på trods af alle disse vanskeligheder kan der entydigt konkluderes, at indekseringskonsistensen er særdeles ringe. Der rapporteres om konsistenser på alt mellem 82% og 4%, men med tyngden lagt i området 30-40%^c.

INDEKSERINGSKONSISTENS OG GENFINDING

I begyndelsen af dette afsnit blev det nævnt, at indekseringskonsistens anses for en væsentlig forudsætning for at dokumenterne kan genfindes. Set i lyset af de resultater, der har kunnet nås på dette område, må situationen siges at være katastrofal.

I den store mængde publikationer, der efter krigen op til 1970 har undersøgt indekseringskonsistens, beskrives relationen mellem indekseringskonsistens og genfindingens kvalitet næsten som en tautologi, eller den antages implicit. Sjældent er forfatteren klar

^a For nærmere gennemgang af de forskellige konsistensmål - bl.a. en del flere end de her nævnte, henvises til Rolling 1981 og Leonard 1977:3-6 og Saracevic 1984.

^b Dette gennemgås i Leonard 1977:7.

^c Markey 1984.

over problemet og foretager en eksplicit antagelse om, at der en sådan sammenhæng.^a

Det er meget forståeligt, at indekserer og andre professionelle biblioteksansatte automatisk forsvarer, at der er en sammenhæng mellem indekseringens konsistens og genfinding af dokumenter. En stor del af deres faglige identitet og uddannelse består i kendskab til og brug af klassifikation og emneord. Derfor er den umiddelbare reaktion, at slutte at indekseringen skal være så konsistent som muligt for brugerens skyld, idet konsistens og kvalitet ofte anses for sammenfaldende. Inkonsistent klassifikation er noget sjusk, der, når det opdages, hurtigst muligt rettes på bibliotekerne.

Ét af de få arbejder, der til dato har søgt at problematisere sammenhængen mellem indekseringskonsistens og genfinding, er Coopers fra 1969^b.

Cooper viste, at det er usikkert, om denne antagelse er teoretisk holdbar.

Hans metode bestod i at opstille antagelser om hvorledes indekseringskonsistens og genfinding hang sammen og dernæst konstruere et modeksempel. Først formulerede han en "stærk hypotese", der lyder:

Hvis metode B producerer en større grad af indekseringskonsistens end metode A, så vil anvendelsen af metode B også resultere i større effektivitet i genfindingen af litteraturen end anvendelse af metode A vil gøre.

Hvordan man end vil præcisere denne hypotese, kan den klart modsiges af det tilfælde, at metode B ganske vist forbedrer konsistensen, men er helt ude af trit med forskningen eller ganske enkelt er noget konsistent sludder.

Dette leder hen til en "svagere" version af ovenstående hypotese, der inddrager konsistensen mellem indekser og bruger ved at antage, at hypotesens gyldighed er betinget af, at bruger-indekser

^a Leonard 1977 og Markey 1986/Interindexer har begge givet gode oversigter over områdets litteratur.

^b Cooper 1969.

konsistensen er mindst lige så stor ved anvendelsen af metode B som ved anvendelse af metode A.

Selv denne svagere version modbevises af Cooper ved at konstruere et modeksempel.

En anden belysning af kompleksiteten i forholdet mellem konsistent emneindeksering og genfindning består i at påvise, at det er muligt at lave søgeprofiler, der giver 100% genfinding af to helt inkonsistente indekseringer, nemlig i de tilfælde søgeprofilen indeholder begge emneindekseringer.

Dokumentbeskrivelse: metode A:	Dokumentbeskrivelse: metode B:	Søgeprofil:
Børnesange	Børnerim	Børnesange eller børnerim

Et mere sandsynligt eksempel er:

Dokumentbeskrivelse: metode A:	Dokumentbeskrivelse: metode B:	Søgeprofil:
Thermoelectricity	Thermoelectric cells	(thermoelectricity eller thermoelectric cells) og (electrons eller elec- tron beams) og (radiation effects eller irradiation)
Electrons	Electron beams	
Irradiation	Radiation effects"	

Analysis

Her har vi igen 100% genfinding ved hjælp af begge beskrivelser, men 0% konsistens mellem disse beskrivelser^a.

Konklusionen er, at vi ikke forstår sammenhængen eller den manglende sammenhæng mellem indekseringskonsistens og genfinding,

^a Dette eksempel er taget fra Rolling 1981:74. Eksemplet belyser yderligere at hvis vi antager, at det er muligt at konstruere den "rigtige" eller "ideale" beskrivelse: "Electron beams, Radiation effects, Thermoelectricity" er genfindingen uafhængig af graden af overensstemmelse med denne beskrivelse - d.v.s. indekseringens absolutte kvalitet - idet de to metoder overlapper med henholdsvis 33% og 66% med den beskrivelse.

hvilket Cooper erklærede allerede i 1969. Emnet er ikke siden taget op på teoretisk niveau.

Disse teoretiske vanskeligheder suppleres af, at det heller ikke er lykkedes at måle om der eksisterer en sammenhæng mellem indekseringskonsistens og genfinding.

Leonard viste ganske vist, at der er en moderat til stærk korrelation mellem indekseringskonsistens og genfinding^a. Undersøgelsen er hverken gentaget eller udviklet siden, og overraskende nok er der ikke foretaget yderligere undersøgelser af denne problematik. De få artikler, der i dag behandler konsistensproblemer^b, antager blot, at der er en sammenhæng mellem indekseringskonsistens og genfinding af dokumenter.

3. OPSUMMERING

Naturlig tekst og kontrollerede sprog har begge deres fordele og ulemper.

Naturlig tekst har nogle fordele i forhold til kontrollerede sprog som værktøj til emneindeksering.

En emnebeskrivelse i naturlig tekst er betydelig mere omfangsrig og redundant, hvilket medfører at beskrivelsen er lettere for læserne at afkode, og derved bliver emnebeskrivelsen mere anvendelig som instrument til at afgøre dokumentets relevans.

En emnebeskrivelse i naturlig tekst beskriver emner tidssvarende, hvilket medfører, at ny banebrydende litteratur kan beskrives adækvat i modsætning til, hvad der er muligt ved anvendelse af de kontrollerede sprog, hvor dagens begreber skal beskrives med fortidens sprog.

^a Leonard 1975, refereret efter Leonard 1977.

^b Ved et opslag i LISA, primo 1994, finder man, at der fra 1970 til i dag gennemsnitligt kun er publiceret ca. 2 artikler om året, der omhandler indekseringskonsistens.

Og endelig tillader emneindeksering med naturlig tekst, at foretage beskrivelser af marginal litteratur, hvis emner, begreber og metoder kan afvige så meget fra de dominerende paradigmer, at de er ubeskrivelige i de dominerende paradigmers sprog, der er det kontrollerede sprog - ordene mangler simpelthen.

Omvendt har emneindeksering med kontrollerede sprog fordele, som emneindeksering med naturlig tekst ikke har.

Den ene fordel består i, at emneindeksering med kontrollerede sprog er en emneindeksering med syntetiserende begreber, hvorved søgeren ikke behøver at finde synonymmer.

Den anden fordel består i, at det ved hjælp af kontrollerede sprog er muligt at foretage entydige emnebeskrivelser.

Og endelig er emnebeskrivelser med kontrollerede sprog uafhængige af det nationalsprog, dokumentet er skrevet på, hvorved søgeren er fri for at oversætte sine søgninger til andre sprog.

Vurderingen af alle disse fordele og ulemper bygger på, at der vælges de rigtige termer.

Det er gennem talrige undersøgelser godtgjort, at konsistent emneindeksering med kontrollerede sprog ikke eksisterer i praksis, og at man heller ikke er i besiddelse af viden af betydning om, hvordan konsistensen kunne øges. Samtidig er det højst usikkert om indekseringskonsistens har nogen som helst indflydelse på, om dokumenterne kan genfindes.

Konklusionen af dette kapitel må derfor være, at der stilles grundlæggende spørgsmål både ved hele vor opfattelse af fordelene og ulemperne ved anvendelse af kontrollerede sprog og naturlig tekst til emneindeksering, og ved vor opfattelse af hvordan emneindeksering overhovedet skal foretages.

Det bliver emnet for næste kapitel.

ANSATSER TIL TEORETISK ANALYSE AF EMNEINDEKSERING

Det er fremgået af de to foregående kapitler, at emneindeksering, emnerepræsentation og emnesøgning er langt fra tilstrækkelig udviklet, og at grundlaget for disse aktiviteter overraskende vakler, set i lyset af vor nuværende viden.

For at få en baggrund til at forstå problemerne og måske foreslå nogle løsninger er det nødvendigt at foretage en nærmere analyse af hele emneindekseringsprocessen og den efterfølgende emneforståelse eller emnerekonstruktion.

Først er det nødvendigt at gøre sig klart hvad man mener med et emne. Hvilken slags begreb er egentlig et dokument's emne? Dette spørgsmål kræver en klargøring ved hjælp af nogle grundlæggende videnskabsteoretiske værktøjer. Det har overraskende nok ganske store konsekvenser for valget af emneindeksering.

Dernæst stiller vi os spørgsmålet: Hvad opfatter brugeren: indekserens emneforståelse, forfatterens emneforståelse, klassifikations- eller emneordssystemets emneforståelse eller måske noget helt andet? Dette analyseres ved anvendelse af begreber fra klassisk modelteori og informationsteori.

Til sidst opregnes konsekvenserne af kapitlets videnskabsteoretiske og informationsteoretiske betragtninger for emneindekseringsprocessen.

1. HVAD ER ET DOKUMENTS EMNE?

Hvad er et dokumentets emne?

Det, der står i titlen, i indholdsfortegnelsen eller i den foretagne emneindeksering? Er det forfatterens emneforståelse eller indeksørens emneforståelse, der er gældende?

Er det, hvad samtiden eller hvad eftertiden mener?

Og hvis der nu er uenighed om, hvad der står i et dokument - hvad er da emnet? Er det så flertallet, der afgør det?

For at afgøre hvad et dokumentets emne er for noget, bør man først gøre sig klart i hvilken forståelsesramme, begrebet hører hjemme. Er et dokumentets emne noget en indeksør bestemmer eller analyserer sig frem til? Konstitueres et dokumentets emne i et informations- og dokumentationsvidenskabeligt felt?

Eller er dokumentets emne allerede dannet i det videnskabelige miljø, hvor dokumentet er produceret?

En sådan begrebsafklaring kræver en diskussion af grundlaget for emnebegreb og emneindeksering set i forhold til de mennesker, der producerer og anvender disse begreber og indekseringer. Derfor skitseres i det følgende nogle videnskabsteoretiske grundbegreber, der i det følgende vil blive brugt til ganske kort at pege på fordele og ulemper ved de behandlede typer emnebegreber og dertil hørende emneindekseringsmetoder.

A. VIDENSKABSTEORETISKE GRUNDBEGREBER

Videnskabsteorien beskæftiger sig med hvordan man producerer videnskab, herunder hvordan man danner begreber og teorier og hvorledes de teoretiske resultater forholder sig til målinger.

To fundamentale områder behandler forholdet mellem det videnskabelige objekt og iagttageren, nemlig ontologien og epistemologien. Ontologien beskæftiger sig med spørgsmålet, om det videnskabelige objekt eksisterer uafhængigt af iagttageren. Epistemologien beskæftiger sig med om det videnskabelige objekt erkendes uafhængigt af iagttageren.

Efterfølgende skitsering af nogle videnskabsteoretiske begreber bygger på den meget udbredte moderne analytiske videnskabsteori, der går under navne som transcendental, kritisk eller videnskabelig realisme^a.

EN REALISTISK ONTOLOGI

Selvom det er en udbredt "plattitude", at iagttageren påvirker resultatet, så er det omvendt helt fundamentalt for enhver videnskab, at iagttageren ikke skaber *alle* de videnskabelige mekanismer, der skal undersøges. I så fald ville objektet for den videnskabelige undersøgelse være iagttageren selv, og videnskaben ville derfor være helt uden interesse for andre. Denne antagelse kaldes ontologisk, hvilket betyder at antagelsen vedrører objektets eksistens. En sådan antagelse kan i og for sig ikke diskuteres.

At antage, at videnskabens objekter og lovmæssigheder er reelt eksisterende, kaldes realisme. Det betyder ikke, at alle ingredienser i en teori er reelt eksisterende.^b Det betyder heller ikke, at en teori kun indeholder observerbare ingredienser.^c Det betyder derimod, at de mekanismer, der søges forklaret, er reelt eksisterende - de er ikke blot tankespind. Det betyder, at det egentlige objekt for videnskabelig forskning hverken er mønstre eller modeller, men de objekter og de mekanismer, der producerer de fænomener, vi iagttager.

EN TRANSCENDENTAL EPISTEMOLOGI

Hvordan erkendes så disse - ontologisk antagne - reelle objekter og mekanismer?

^a Den realistiske videnskabsteori, der skal refereres til her, er den, der er udviklet af Roy Bhaskar m.fl. i 1970'erne under betegnelsen transcendental realisme, se f.eks. Bhaskar 1975.

^b Selv i fysikken er der adskillige begreber, der ikke kan siges at eksistere, men tjener som et led i en teori eller en model, til at forklare reelt eksisterende fænomener f.eks. ved udveksling af virtuelle partikler i stødprocesser.

^c Også her byder fysikken på adskillige ikke-målelige begreber, for eksempel noget så dagligdags som masse. Hvad man måler er afledede størrelser, der er målelige.

I lang tid har videnskabsteoretikere hævdet, at erkendelsen af lovmæssigheder blev vundet alene ved at iagttage fænomenerne. Dette er imidlertid langt fra tilfældet, hvilket praktiserende videnskabsfolk da også vidste.

Tilfældige korrelationer som et fald i antallet af børnefødsler og storke i Sydsverige over en ti-årig periode viser, at der ikke behøver at være *reelle* sammenhænge mellem data. Og omvendt afvises geologiske teorier ikke, bare fordi en boring undlader at bekræfte en teoretisk forudsigelse af en olieforekomst - man vil oftest sige, at der var andre faktorer, der spillede ind i det konkrete tilfælde.

Kun i såkaldt "lukkede systemer", hvor man kan isolere de mekanismer, der ønskes undersøgt, kan videnskabelige teorier have muligheder for at forudsige fænomener^a. Dette er hvad der tilstræbes i laboratorier, der på denne måde kan siges at være en til formålet konstrueret empiri.

Men virkeligheden er normalt ikke lukket. Den er åben for et utal af "teorifremmede" påvirkninger.

Derfor kan videnskaben normalt ikke forudsige, men kun forklare fænomener i naturen, mellem mennesker eller i mennesker.

Hvis man herefter udelukker muligheden af, at de ontologisk antagne mekanismer og objekter erkendes religiøst eller genetisk, må videnskabelig viden produceres af mennesket. Det medfører, at videnskabelig viden også er social og historisk.

Det medfører på den anden side ikke, at den videnskabelige viden *kun* er social og historisk, da den videnskabelige produktion netop er bundet til at forklare de faktiske fænomener.

Transcendental realisme opfatter altså videnskabernes mekanismer og objekter som værende reelle, virkeligt eksisterende, uafhængige af mennesket. Den videnskabelige erkendelse, derimod, består i at konstruere en social og historisk form i vores hoveder med det empiriske udgangspunkt, der foreligger i fremtrædelsesformerne.

^a Se Bhaskar 1975: Kapitel 2 om "Actualism and the Concept of Closure".

B. DET INFORMATIONS- OG DOKUMENTATIONSVIDENSKABELIGE EMNEBEGREB

I den informations- og dokumentationsvidenskabelige litteratur opfattes begrebet emne som grundlæggende bundet til dokumentet på den ene side og brugeren af informationssystemet på den anden side. De tilsvarende indekseringsmetoder kaldes dokument- eller indholdsorienterede h.h.v. behovs- eller spørgsmålsorienterede.^a

Anvendes en indholdsorienteret indekseringsmetode, foretager indeksøren en tekstanalyse af dokumentet. I analysen betragtes teksten som en kommunikation mellem forfatter og læser, eksempelvis af spørgsmålet: Hvad mener forfatteren?

Men teksten kan også analyseres ud fra f. eks. ideologikritiske, sociologiske, psykoanalytiske metoder, alle med det fælles udgangspunkt, at et dokument kan tolkes på mange måder - også måder, forfatteren ikke selv var klar over eller ville acceptere^b.

En sådan analyse vil selvfølgelig være afhængig af den indeksør, der foretager den.

Anvendes den behovsorienterede indekseringsmetode skal indeksøren danne sig et billede af den fremtidige brugers behov for det aktuelle dokument. Dette gøres ofte i forhold til bestemte målgrupper, der i særlig grad betjenes med det pågældende informationssystem.

Helt fundamentalt står indeksøren i en gætte-situation, hvor det er indeksørens opfattelse af de fremtidige behov hos brugerne, der skal indekseres efter.

Lige meget hvor snedigt man kvalificerer og beskriver dette gætteri, dets formål og interesser, er og bliver det gætteri, der medfører, at det er forskelligt fra indeksør til indeksør.

^a Se f.eks. Lancaster 1991:8-9.

^b Under disse omstændigheder tæller forfatterens tolkning som én blandt mange mulige tolkninger. Og der er god grund til at påpege, at anvender man forfatterens tekst som beskrivelse af dokumentets emne, er man offer for den fejltagelse, der tager det umiddelbart iagttagelige som det videnskabelige objekt, mens der i virkeligheden er tale om at majorisere én ud af mange mulige subjektive erkendelser.

Indenfor samme traditioner har Birger Hjørland bidraget med et emnebegreb. Han definerer et dokumentets emne som dokumentets erkendelsesmæssige potentialer, formuleret og analyseret indenfor de faglige miljøer, dokumentet er den del af^a.

Dette emnebegreb hævdes at være objektivt.

"Når vi fastholder, at dokumentets egenskaber er objektive, selv om beskrivelsen af dem kræver særlige subjektive forudsætninger, så ligger der i dette, at virkeligheden, afprøvningen af dokumentet i praksis, til syvende og sidst vil afgøre dets erkendelsespotentialer uanset hvor mange misopfattelser, der forinden har gjort sig gældende. Det er således historien, der i sidste instans er den endelige dommer vedrørende objektiviteten af udsagn om dokumentets egenskaber. Og selv om historien naturligvis sjældent vil kunne afgøre dette endeligt, fastholder vi begrebet om objektive egenskaber ved dokumenter, der ligger til grund for vore bestræbelser med at analysere dem."^b

Denne ontologi er en slags futuristisk realisme, idet det hævdes, at det, vi studerer eller søger at analysere, er reelt eksisterende - ikke noget vi kun har i vore hoveder, men noget, der viser sig i fremtiden.

Ser vi på erkendelsesmåden eller epistemologien er den subjektiv. Dette understreges senere:

"En emnebeskrivelse af et dokument er således på den ene eller anden måde et udtryk for dokumentets erkendelsesmæssige potentialer, således som de fremtræder for den, der foretager emnebeskrivelsen. Jo bedre beskrivelsen forudsiger dokumentets potentialer, jo vigtigere, mere objektiv, er emnebeskrivelsen. Omvendt må en tolkning af en given emnebeskrivelse medinddrage forudsætningerne hos den person, der har foretaget emnebeskrivelsen... Hvis man skal sige, hvad man kan forvente at finde under en given position i et emnedatasystem, så må man foretage en rekonstruktion. Først må man rekonstruere, hvordan emneopfattelsen har været hos dem, der har foretaget emneanalysen."^c

^a Hjørland 1993:120.

^b Hjørland 1993:116-117.

^c Hjørland 1993:120.

En nødvendig betingelse for at kunne rekonstruere emnebeskrivelsen er altså, at man kender emneopfattelsen hos den, der har foretaget emnebeskrivelsen, eller at emnet er bredt og alment accepteret i vor kulturkreds.

C. DET FAGVIDENSKABELIGE EMNEBEGREB

I det aktivt producerende videnskabelige miljø arbejder man med et andet emnebegreb. Et emne konstitueres af og er en del af det faglige paradigme, traditioner, objektet, metoderne og teorien. Således er der tale om, at emnebegrebet er en del af det videnskabelige kollektiv, og er med til at definere dette. Når der rapporteres fra det videnskabelige miljø, ligger der altså et reelt eksisterende emne bag dokumentet.

Denne opfattelse medfører at begrebet emne strengt taget ikke har mening, brugt som et dokumentets emne, men som arbejdets emne. Et dokument betragtes som en fremtrædelsesform for det reelt - og objektivt - eksisterende emne i det bagved liggende videnskabelige arbejde.

Ved at læse eller analysere et videnskabeligt dokument kan man så søge at producere et billede af det bagved liggende videnskabelige arbejdes emne. Det er forfatterens formulering af emnet, der er den autoritative kilde til viden om det videnskabelige arbejde, forfatteren selv har udført. Forfatterens formuleringer er en kommunikationskanal for det videnskabelige arbejde - så at sige et led i Kooperationen mellem videnskabsfolk. I modsætning til, hvad der var tilfældet ved anvendelse af det informations- og dokumentationsvidenskabelige emnebegreb, repræsenterer forfatterens egen emneopfattelse *alene* dokumentets emne^a.

Dette udelukker ikke, at nogle forfattere kan være dårlige formidlere af deres emne. De kan postulere, at et dokument handler om et emne, der kun berøres marginalt i pågældende dokument, eller de kan have misforstået emnet. Men det er ikke en anonym indeksørs

^a Sammenlign fodnote b på side 53

opgave at rette op på dette forhold. Det er det videnskabelige samfunds og de faglige miljøers historisk specifikke opgave at bedømme dette.

Det fagvidenskabelige emnebegrebs ontologi er ikke afhængigt af nogen indeksør eller bruger.

D. ANVENDELSE AF DE TO EMNEBEGREBER

Vi har set, at der findes to grundlæggende forskellige emnebegreber: Det informations- og dokumentationsvidenskabelige emnebegreb og det fagvidenskabelige emnebegreb.

DET INFORMATIONS- OG DOKUMENTATIONSVIDENSKABELIGE EMNEBEGREB

Det informations- og dokumentationsvidenskabelige emnebegreb er grundlæggende for klassifikation og udarbejdelse af bibliografiske værker - f.eks. oversigtsværker og egentlige bibliografier. Vilkårene for anvendelse af dette emnebegreb er, at indekseringerne er subjektive, og derfor kun anvendelige til rekonstruktion af dokumentets emne, hvis man får tilstrækkelig information til at kunne rekonstruere indeksørens emneopfattelse eller hvis der kun eksisterer ét perspektiv på vedkommende emne.

Med denne argumentation bliver anvendelse af klassifikation og emneord til emnebeskrivelse kun hensigtsmæssig, brugt på begreber eller emner, der er alment accepterede. Tilbage er de bibliografiske værker og oversigtslitteraturen, der normalt har en tilstrækkelig fyldig beskrivelse af emnerne i den beskrevne litteratur til, at det er muligt at danne sig et billede af indeksørens emneopfattelse.

DET FAGVIDENSKABELIGE EMNEBEGREB

Det fagvidenskabelige emnebegreb er grundlæggende for anvendelse af indholdsfortegnelse, abstract eller andre dele af dokumentets tekst som emneindeksering. Ved brug af en sådan indeksering er det muligt at rekonstruere det videnskabelige arbejdes emne.

I den udstrækning indeksøren selv vælger ud af dokumentets tekst ved emneindekseringen, tilføjes indekseringen en subjektiv faktor, nemlig udvælgelsen af det, indeksøren skønner som væsentligt. Denne subjektivitet er det muligt at reducere ved at konstruere

objektive udvælgelseskriterier. Eksempelvis kan man kræve at kun dele af indholdsfortegnelsen eller stikordsregistret, der refererer til mere end et vist antal sider medtages, som det gøres i SAP-indeksering^a, eller at hele indholdsfortegnelsen altid kommer med, som det gøres i de fleste analytiske kataloger^b og i PARADOKS-indeksering.

Sammenlignet med den subjektivitet, der nødvendigvis indgår i enhver anvendelse af det informations- og dokumentationsvidenskabelige emnebegreb, er enhver anvendelse af selve dokumentets tekst dog en minimering af subjektiviteten i emnebeskrivelsen, da forfatterens ordvalg er bibeholdt. Men grundlæggende repræsenterer udvælgelsen af dele af dokumentets tekst til brug for emneindeksering naturligvis et subjektivt element.

E. KONKLUSION

I det foregående har vi set på nogle konsekvenser af at anvende det informations- og dokumentationsvidenskabelige emnebegreb og det fagvidenskabelige emnebegreb. På det foreliggende grundlag kan der konkluderes, at den mest brugte emneindeksering i bibliotekskataloger, klassifikation- og emneordstildeling, er uhensigtsmæssig, da det ofte ikke er muligt at rekonstruere emnet ud fra indekseringen. Tilbage står vi med to anvendelige former for emneindeksering, der hviler på hvert sit emnebegreb: Den bibliografiske publikation og udvalg af tekst fra selve dokumentet.

Ingen af de to former for emneindeksering er fuldkomne, men de repræsenterer begge en betydelig kvalitetsforbedring i forhold til klassifikations- og emneordsindeksering.

^a Se beskrivelsen af SAP-indekseringen afsnit 5 side 90.

^b Se afsnit 5.1 side 92 om analytiske kataloger.

2. TOLKNINGER, MODELLER OG EMNER I KLASSISK MODELTEORI

Når en forfatter skriver et dokument eller udarbejder et emnedatasystem er det på baggrund af viden, begreber, teorier og holdninger, der er socialt og historisk bestemt. Ideen til dokumentet overstiger altid selve dokumentet, og når dokumentet først er skrevet repræsenterer forfatterens idé med dokumentet én af mange mulige tolkninger af dette dokument. Det samme gælder for et emnedatasystem. Denne kommunikationsproces kræver, som enhver anden kommunikationsproces, at der er nogle fælles referencerammer for forfatter og læser. Ellers er læserens udbytte så fjernt fra forfatterens intentioner, at der i alle tilfælde ikke er tale om en emnebeskrivelse fra forfatter til læser i sædvanlig forstand.

Men kunne man ikke opnå en entydig kommunikation eller blot opstille et begrænset antal muligheder for tolkninger?

Kunne forfatteren ikke - ved at gøre sig umage - sørge for, at dokumentet eller emnedatasystemet kun kan fortolkes på én måde?

Dette spørgsmål søger vi at besvare i det følgende.

En sådan besvarelse kræver en præcision af begrebet "tolkning", "det system, der tolkes" samt en præcision af disse to begrebsrelationer.

Da så komplekse størrelser som dokumenter og emnedatasystemer ikke kan undersøges generelt, vil vi i det følgende begå en "reduktionisme": Vi beskæftiger os med simple eller enkle systemer og deres tolkning eller, som det oftest hedder, modeller. Hvis det viser sig, at ganske simple systemer som geometriske eller kemiske systemer altid kan tolkes på et utal af måder, så antager vi, at det også gælder for mere komplekse systemer som dokumenter og emnedatasystemer. Og hvis vi finder forhold i de simple systemer, der øger eller begrænser antallet af tolkninger, så antages det, at det samme gælder for de mere komplekse systemer.

For at vise hvorledes tolkningen af - eller modellen for - et system er afhængigt af systemets indhold, gennemgås først et simpelt

geometrisk eksempel. Herunder vises eksempler på ugyldige tolkninger og hvordan disse ugyldige tolkninger kan gøres gyldige og udgøre en egentlig model for det teoretiske system. Dernæst betragtes udviklingen af det kemiske affinitetsbegreb i perioden omkring Dalton. Her træffer vi på et teoretisk system, der er utilstrækkeligt, eller tolkninger, der ikke er teoretisk belæg for.

Med udgangspunkt i de gennemgåede eksempler behandles til sidst den formelle logiks behandling af begreberne syntaktisk system og semantisk struktur, herunder begrebet model, for at kunne præsentere nogle af de resultater, der er relevante for dette afsnits problemstilling:

Kan man tolke dokumenter eller emnedatasystemer på én eller blot et begrænset antal måder?

Hvad påvirker antallet af tolkningsmuligheder og hvorfor?

A. GEOMETRI - ET EKSEMPEL TIL BELYSNING AF MODELLERS AFHÆNGIGHED AF AKSIOMERNE I DET TEORETISKE SYSTEM

At finde en model for et formelt system er det samme som at finde en tolkning - noget for hvilket alle aksiomerne i det formelle system er sandt.

Euklids rumlige geometri er et eksempel på et formelt system. Det formelle system består bl.a. af aksiomer som det velkendte: "Gennem et punkt udenfor et plan kan lægges et og kun et plan parallelt med det givne".

En model skal også angive, hvad et punkt er, og hvad et plan er. Som aksiomerne står, er de nemlig bare påstande om nogle udefinerede størrelser, og således hverken sande eller falske.

Modellen for Euklids geometris formelle system er det sædvanlige tredimensionale rum, der da også kaldes "det euklidiske rum". I dette rum eller model kender vi aksiomets mening og ved, at det er sandt.

Men er en kugleoverflade i det sædvanlige 3-dimensionerede rum også en model for Euklids geometri?

Hvis "et punkt" i det formelle system svarer til et punkt på kugleoverfladen og "et plan" i det formelle system svarer til "en storcirkel" oversættes tolkningen af aksiomet

"Gennem et punkt uden for et plan kan lægges et og kun et plan parallelt med det givne plan"

til

"Gennem et punkt (på kuglen) uden for en storcirkel kan lægges en og kun en storcirkel parallel med den givne storcirkel".

Da der ikke findes parallelle storcirkler på en kugleoverflade - tværtimod skærer alle storcirkler hinanden i to punkter - er aksiomet vurderet som værende falsk i denne tolkning.

Eller: en sædvanlig 3-dimensional kugleoverflade, hvor hvert "punkt" i Euklids rumlige geometris formelle system repræsenteres af et punkt på kugleoverfladen, og hvert "plan" repræsenteres af en storcirkel, er ikke en model for Euklids rumlige geometris formelle system.

Derimod er en 3-dimensional kugleflade, hvor hvert "punkt" svarer til et punkt på kugleoverfladen, og "plan" svarer til en vilkårlig lille- eller storcirkel en model for Euklids rumlige geometris formelle system. Dette stemmer helt overens med vor dagligdags forståelse af kuglefladers geometri.

Hvis vi nu ændrer det omtalte aksiom i det formelle system til

"Gennem et punkt udenfor et plan kan ikke trækkes noget plan parallelt med det givne",

men fastholder de øvrige aksiomer, er det formelle system ikke mere Euklids geometri. Dette formelle system kaldes Riemann's geometri. Og en model for Riemann's geometri er beskrevet ovenfor: en sædvanlig 3-dimensional kugleoverflade, hvor hvert "punkt" svarer til et punkt på kugleoverfladen, og "plan" svarer til storcirkel. Den sfæriske geometris sætninger er netop sætninger

udtrykt i modellens sprog, f.eks. er summen af vinklerne i en trekant er større end 180° , hvilket er falsk i det Euklidiske rum.

Aksiomet kan også ændres til

"Gennem et punkt uden for et plan kan føres flere planer, parallelle med det givne plan".

Dette formelle system kaldes Lobatchevsky's geometri, hvor summen af vinklerne i en trekant altid er mindre end 180° .

Det fremgår tydeligt, at aksiomerne udvælger de mulige modeller, og at en ændring af et aksiom giver mulighed for nye modeller. Der er imidlertid visse bånd på, hvilke aksiomer der kan sammenstilles, hvis det formelle system skal have en model - indbyrdes modstridende aksiomer tillader ingen model.

B. ELEMENTER AF KEMIENS UDVIKLING OMKRING DALTON

Gennem meget af det attende og ind i det nittende århundrede troede næsten alle europæiske kemikere, at alle kemiske stoffer bestod af elementære atomer, holdt sammen af gensidigt tiltrækkende kræfter. Således blev en klump sølv holdt sammen af sølv atomernes indbyrdes affinitet. Samme teori forklarede, at sølv opløses i syre, eller salt i vand, fordi sølv partiklers affinitet over for syrepartikler, eller saltpartiklers affinitet over for vandpartikler, var stærkere end den indre affinitet hos de indgåede stoffer.

Datidens kemikere anvendte en sådan affinitetsteori med stor effekt f.eks. til en inddeling af stofferne i forskellige kategorier, deres art (syre, salt) og efter deres styrke, hvorved kemiske reaktioner kunne forudsiges.^a

I det følgende skal vi se på affinitetsteoriens skelnen mellem fysiske blandinger og kemiske forbindelser.

^a Se f. eks. Metzger 1930:142.

Affinitetsteorien skelner mellem fysiske blandinger og kemiske forbindelser gennem følgende antagelse:

En kemisk forbindelse er en forbindelse, hvor de indgående stoffer har tilstrækkelig stor relativ affinitet.

Det teoretiske begreb "tilstrækkelig stor relativ affinitet" kan omformuleres til kemisk affinitet, A_k . Det teoretiske begreb A_k opdeler da den teoretiske models stoffer i to mængder: de der har kemisk affinitet (kemiske forbindelser) og de der ikke har kemisk affinitet (fysiske blandinger).

Denne teori kan ikke umiddelbart forbindes til eksperimenter, da affinitetsbegrebet ikke kan forbindes til en målelig størrelse. For at teorien skal blive anvendelig, må den indeholde mindst et teoretisk begreb, der kan forbindes til en målelig størrelse.

Datidens kemikere antog da, at der blev dannet en kemisk forbindelse, når reaktionen producerede lys, luftbobler, varme m.v., mens en ikke-kemisk forbindelse var inhomogen.

Dette teoretiske system var i nogle tilfælde ganske ubrugeligt. Således blev tilfælde som salt og vand, legeringer og atmosfærisk luft slet ikke "dækket" af teorien idet disse blandinger hverken dannes samtidig med produktion af lys, varme, luftbobler etc. eller er inhomogene.

De fleste kemikere på den tid ændrede da deres teoretiske fundament, således at en kemisk forbindelse i stedet for at være en forbindelse dannet sammen med produktion af lys, varme etc., nu blot var en homogen forbindelse.

KRISE

I slutningen af det attende århundrede var det alment kendt, at visse kemiske forbindelser havde konstant vægtforhold og andre havde det ikke. Utilstrækkeligheden af teorien viste sig ved, at teorien ikke kunne gøre rede for denne opsplitning.

Der gik en lang periode, hvor disse anomalier ikke indgik med nogen vægt i den kemiske videnskab.

KRISEN OPHÆVES

Krisen ophævedes ved "en videnskabelig revolution" - ved en grundlæggende omstrukturering af den kemiske videnskab.

Selve det grundlæggende begreb, kemisk forbindelse, blev ændret ved konstruktion af en helt ny teori: Dalton's kemiske atomteori.

Dalton var meteorolog og havde aldrig før været interesseret i kemi. Som meteorolog beskæftigede han sig med det fysiske problem: absorption af vand i atmosfæren. Han betragtede de homogene blandinger af luftarter og vanddampe som fysiske blandinger - styret af fysiske processer. Derfor var blandingerens ensartethed (homogenitet) noget, han troede kunne løses ved kendskab til de forskellige atompartiklers størrelse og vægt. Det var i forsøget på at bestemme disse størrelser og vægte, at Dalton beskæftigede sig med kemi. Han antog fra begyndelsen, at i de få reaktioner, han anså som kemiske (som fysiker anså han det meste som fysisk, ligesom kemikerne før havde anset det meste for at være kemisk), kunne atomerne kun kombineres en og en eller i et forhold svarende til forholdet mellem to hele tal. Denne for ham naturlige opfattelse satte ham i stand til at bestemme størrelse og vægt af de atomer, der indgik i en kemisk reaktion og selvfølgelig - som et "biprodukt" - fulgte også "loven om de konstante vægtforhold", der forklarede de iagttagne konstante vægtforhold i kemiske forbindelser, som var årsag - eller én af årsagerne - til krisen.

C. EMNEDATASYSTEMER

Emnedatasystemer opdeles sædvanligvis i to hovedtyper, nemlig hierarkiske klassifikationssystemer og thesauri.

KLASSIFIKATIONSSYSTEMER

Et hierarkisk klassifikationssystem er et system, der grundlæggende bygger på partitive relationer mellem klasserne - dvs. et emne kan opdeles i flere underemner. Så længe de hierarkiske systemer består af sådanne relationer, er systemet stabilt og kan redegøre for de begreber, der ligger indenfor systemets egen forståelsesramme.

Til trods for, at alle klassifikationssystemer har en vægtig indbygget konservatisme - det er faktisk deres styrke, som vi skal se - ændres

de, efterhånden som de videnskabelige arbejdsområder og begreber ændrer sig. Med adskillige års forsinkelse foretages der enten ad hoc revisioner eller der publiceres helt nye udgaver af klassifikationssystemerne. Ofte består sådanne ændringer i, at der tilføjes nye tværgående emner, hvor det ikke er meningen, at relationerne skal være partitive, fordi emnerne ikke passer ind i systemets oprindelige valg af "grundlæggende emner".

EKSEMPEL

Hvor skal forskningen i kunstig intelligens anbringes - under lægevidenskab/neurologi, under psykologi eller under datalogi? I RUB's lokale variant af UDK er kunstig intelligens anbragt som *én* gruppe under datalogi med en kolonforbindelse til psykologi (681.3:159.9).

Hermed er der indført en i streng forstand partitiv relation, der definerer al forskning i kunstig intelligens som en del af datalogien. Det vil klart være ude af trit med virkeligheden, hvorfor man i sådanne tilfælde opgiver at opfatte relationerne som partitive, hvilket ellers var systemets grundidé.

THESAURI

Mens klassifikationssystemer i princippet bygger på en stærk indre struktur, der ganske vist "smuldrer ved brug", så bygger thesauri på en meget pragmatisk fremgangsmåde. Dette afspejler sig tydeligt i ISO-standardens definition:

En thesaurus defineres "funktionelt som et terminologisk kontrolinstrument, der anvendes, når det naturlige sprog i dokumenter ved indeksering eller brug omsættes til præcist afgrænsede termer (dokumentationssprog...)"^a.

Et par gode eksempler på denne type emnedatasystemer finder vi i EUDISED thesaurussen, hvor relationer af typen narrower/-broader/related terms er eksemplificeret nedenfor^b.

^a ISO-standard 2788.

^b Viets & van Slype 1984:193 og 293.

Videnskab og teknologi	Offentlig forvaltning
Naturvidenskaber	Retfærdighed
Fysiske Videnskaber	Dom
Astronomi	Sanktion
Fysik	Frihedsstraf
Akustik	Fange
Audiometri	Domstol
Optik	Myndighedsområde
Geovidenskaber	

Mens relationerne mellem "Videnskab og teknologi" og "Naturvidenskaber", mellem "Naturvidenskaber" og "Fysiske Videnskaber", mellem "Fysiske Videnskaber" og "Fysik", mellem "Fysik" og "Akustik" er partitiv, er dette ikke tilfældet for relationen mellem "Akustik" og "Audiometri", idet "Audiometri" er læren om at måle ørets modtagelighed over for lydindtryk - altså nærmest et medicinsk emne.

Tilsvarende ses det, at der er partitive relationer mellem hele rækken af begreber fra offentlig forvaltning til frihedsstraf, mens begrebet "fange" ikke er et underbegreb til frihedsstraf i sædvanlig sprogbrug.

Vi ser, at systemforskellen mellem de hierarkiske klassifikationssystemer og thesauri mere er et spørgsmål om oprindelse end egenskaber i forbindelse med praktisk brug. Begge systemer må nødvendigvis operere som ufuldstændige systemer, der grundlæggende er bundet til vores sædvanlige forståelse af ordenes betydning. Det betyder, at såvel indeksøren som brugeren af emnedatasystemet i vid udstrækning er henvist til deres egne opfattelser af hvilke begreber, der ligger bag de opskrevne ord.

D. DET VIDENSKABELIGE MODELBEGREB

I det foregående har begreberne system og model været anvendt gennem konkrete eksempler, og det er blevet klart, at systemets forhold til modellen er et centralt spørgsmål for enhver videnskab.

Systemer er anvendt i betydningen teoretiske, formelle systemer. Det teoretiske system var en række sammenhænge mellem teoretiske begreber.

Modeller er anvendt i betydningen løsningssættet til eller tolkningen af ovennævnte systemer.

Ved gennemgangen af de konkrete eksempler er der allerede flere gange forekommet problemer, hvor en nærmere undersøgelse af modellens og systemers egenskaber og forhold til hinanden er påkrævet.

For eksempel var den teoretiske kemi på et tidspunkt i den stabile periode før Dalton uanvendelig, fordi den klassificerede kemiske forbindelser som forbindelser, der dannedes ved produktion af lys, varme etc., og fysiske blandinger som inhomogene tilstande. Men hvad så med salt og vand eller atmosfærisk luft? Begge dele er homogene uden at være dannet ved produktion af lys eller varme. Problemet er af en meget generel natur; det teoretiske system er ikke "godt nok", eller emnedatasystemet kan ikke følge med virkeligheden.

For at kunne bearbejde ovennævnte problemer må der foretages en præcisering af, hvad der menes med et system og en model.

En sådan præcisering vil forme sig som en opskrivning af systemets (syntaktiske) grundbegreber, der eksemplificeres ved en anvendelse på kemiens udvikling omkring Dalton efterfulgt af en opskrivning af de semantiske grundbegreber, hvorunder begrebet model konstrueres.

SYNTAKTISKE GRUNDBEGREBER

Et formelt eller teoretisk system er ikke andet end en formel maskine, der laver nye formler og udsagn af de oprindelige formler og udsagn efter visse regler. Der optræder ingen vurderinger af formlers og udsagns sandhedsværdi.

Et teoretisk system består af:

1. Nogle symbolske objekter, der inden for syntaksen er de egentlige objekter:

$$a_1, a_2, a_3, \dots$$

$$b_1, b_2, b_3, \dots$$

$$c_1, c_2, c_3, \dots$$

altså en uendelig, men tællelig, række bogstaver. Disse objekter kaldes individkonstanter.

2. Nogle egenskaber, ved hvis hjælp man kan danne nogle elementære udsagn om individkonstanter (objekterne):

$$P_1, P_2, P_3, \dots$$

$$Q_1, Q_2, Q_3, \dots$$

$$R_1, R_2, R_3, \dots$$

hvor $P_2(c_3)$ betyder at c_3 har egenskaben P_2 . Disse egenskaber kaldes prædikatkonstanter.

3. Nogle individvariable, ved hvis hjælp man kan angive, at et udsagn ikke kun skal tages for en kendt prædikatkonstant:

$$x_1, x_2, x_3, \dots$$

$$y_1, y_2, y_3, \dots$$

$$z_1, z_2, z_3, \dots$$

4. Nogle kvantifikatorer til brug for udsagn, hvori individvariable indgår:

$\exists x: P(x)$ læses: der eksisterer et x således, at x har egenskaben P . \exists kaldes eksistenskvantifikatoren.

$\forall x: P(x)$ læses: for alle x gælder det, at x har egenskaben P . \forall kaldes universalkvantifikatoren.

5. Nogle måder at kombinere elementære udsagn på, f.eks.:

implikation: \rightarrow og negation: $-$

6. Nogle dannelsesregler for sammensætning og skrivning af symboler - f.eks:

$P(a)$, men ikke $a(P)$

7. Nogle deduktionsregler, slutningsregler, f.eks.:

hvis $A \rightarrow B$ og A er en del af systemet, så følger B (separationsreglen)

8. Nogle aksiomer, der udgør det teoretiske indhold.

Alt dette er i og for sig trivielt viden, som det bare er lidt usædvanligt at se eksplicit nedskrevet. Alle anvender det imidlertid - mere eller mindre fejlfrit - uden at tænke nærmere over det. Det eneste, der er grund til at hæfte sig ved, er det formelle systems strenge - maskinagtige karakter - ord som sand, falsk, rimelig eller virkelighed forekommer ikke. Systemet er bare til, for at udlede nogle teoremer fra nogle aksiomer.

Som eksempel på syntaksens formalisering skal vi igen se på kemiens udvikling omkring Dalton.

I den stabile periode før Dalton skelnede man i teorien mellem kemiske forbindelser og fysiske blandinger ved at en kemisk forbindelse havde tilstrækkelig stor affinitet, A_k , mellem de indgående stoffer til at der ved dannelsen blev produceret lys, varme m.v., mens en fysisk blanding var inhomogen. Dette teoretiske system kan opskrives som følger:

a (individkonstanter)

x (individvariable)

Prædikatkonstanter:

A_k (kemisk affinitet)

P (produktion af lys etc.)

H (homogenitet)

Aksiomer

$\forall x: A_k(x) \rightarrow P(x)$

$\forall x: \neg A_k(x) \rightarrow \neg H(x)$

(1)

Systemet har den mangel, ikke at udtale sig om homogene forbindelser. En saltvandsopløsning eller en legering kan ikke beskrives, da de hverken udsender lys etc. eller er inhomogene.

Formelt kan manglen udtrykkes ved at hverken $A_k \rightarrow H$ eller $\neg(A_k \rightarrow H)$ kan udledes af systemet. En sådan mangel har et navn: Systemet siges at være umættet. Dette har - som det bliver gennemgået i afsnittet om semantiske grundbegreber - konsekvenser for den til systemet svarende model.

Omvendt gælder det at et system er mættet, hvis det gælder for ethvert udsagn A , dannet af systemets begreber, at enten A eller A 's negation ($\neg A$) kan udledes af det formelle system.

Datidens kemikere ændrede senere deres begreb om kemisk forbindelse til at være en homogen forbindelse slet og ret. Det betød at saltvandsopløsninger, legeringer og atmosfærisk luft blev betragtet som kemiske forbindelser. Så ved at ændre aksiomerne

$$\begin{aligned}\forall x: A_k(x) &\rightarrow P(x) \\ \forall x: \neg A_k(x) &\rightarrow \neg H(x)\end{aligned}$$

i ovenstående teoretiske system til

$$\begin{aligned}\forall x: A_k(x) &\rightarrow H(x) \\ \forall x: \neg A_k(x) &\rightarrow \neg H(x)\end{aligned}\tag{2}$$

blev systemet mættet.

SYNTAKTISKE NORMER

Ovenfor blev begrebet et mættet system nævnt. Mætning er et eksempel på en egenskab ved nogle syntaktiske systemer - det man kalder en syntaktisk norm.

I modelteorien arbejder man også med en del andre normer. En syntaktisk norm er systemets indre mening. Hvis f.eks. et system indeholder alle tænkelige sammenhænge mellem systemets teoretiske begreber, er det meningsløst.

KONSISTENS

Modelteoriens konsistensbegreb er en syntaktisk norm:

Et formelt system er konsistent, hvis blot én sammenhæng mellem systemets begreber ikke kan udledes af systemet.

I mere dagligt sprog opererer man med en syntaktisk norm, der også kaldes konsistens:

Et formelt system er konsistent hvis ikke både formlen A og dens negation $\neg A$ kan udledes af systemet.

Det kan vises at de to konsistensbegreber er ækvivalente^a.

MÆTNING

Og endelig skal den syntaktiske norm mætning gentages her, da den vil være af betydning for vores konklusioner vedrørende tolkninger af dokumenter og emnedatasystemer:

Et formelt system er mættet, hvis det gælder for ethvert udsagn, A , dannet af systemets begreber, at enten A eller $\neg A$ kan udledes af det formelle system.

Mætning betyder altså, at systemet er tilstrækkeligt udbygget til at bearbejde hele det område, der er givet ved samtlige kombinationer af systemets begreber.

Det er yderst sjældent - og for øvrigt svært at afgøre - at systemer er mættede.

Kurt Gödel har for aritmetikkens vedkommende konstrueret et selvrefererende^b udtryk af aritmetikkens formelle system, hvor hverken formelen selv eller dens negation kunne udledes af systemet. Han har altså vist, at aritmetikkens formelle system er syntaktisk umættet.

SEMANTISKE GRUNDBEGREBER

Der skal nu etableres en løsning til eller en tolkning af det syntaktiske system. Der skal etableres en "ydre mening" med det. For første gang indføres begreberne "sand" og "falsk".

^a Se f.eks. Badiou 1969:appendix 8.

^b Et selvrefererende udtryk er eksempelvis det fuldstændige detaljerede kort over verden. Det leder til et paradoks, da dette kort, ifølge forudsætningen, må indeholde et detaljeret kort over kortet over verden o.s.v.. Et andet eksempel er det klassiske, at vi definerer en barbér, som den, der klipper de andre. Hvem klipper så barberen?

Igen skal kemiens udvikling omkring Dalton tjene som et introducerende eksempel.

Det teoretiske system (1):

a (individkonstanter)

x (individvariable)

Prædikatkonstanter

A_k (kemisk affinitet)

P (produktion af lys etc.)

H (homogenitet)

Aksiomer

$\forall x: A_k(x) \rightarrow P(x)$

$\forall x: \neg A_k(x) \rightarrow \neg H(x)$

kan tolkes på følgende måde:

Lad der være givet en mængde af sammensatte stoffer v_1, v_2, v_3, \dots . Disse sammensatte stoffer danner mængden V .

Lad der være givet en gruppe af delmængder af V : $[a_k V]$, $[pV]$ og $[hV]$, svarende til de i det teoretiske system forekommende prædikatkonstanter.

Hvis denne struktur skal være en model for det teoretiske system skal de i V indgående kemiske forbindelser producere lys, varme eller lignende ved deres dannelse, og de ikke-kemiske forbindelser (blandinger) skal være inhomogene. Dette kan formuleres som:

$$\begin{aligned} [a_k V] &\subset [pV] \text{ og} \\ [-a_k V] &\subset [-hV] \end{aligned} \quad (3)$$

hvor $a_k V$ er forbindelser med kemisk affinitet, pV er forbindelser, dannet sammen med varme, lys, gasser m.v., og $-hV$ er inhomogene forbindelser.

Af (3) følger at $[-pV] \cap [hV]$ er en tom mængde, hvilket betyder, at modellen ikke behandler homogene tilstande, der ikke udsender

lys, varme etc.. Det betyder at salt i vand eller atmosfærisk luft ikke er beskriveligt i den opstillede model. Denne mangelfuldhed skyldes som nævnt, at det teoretiske system er umættet.

Ved at erstatte P med H i aksiomerne ovenfor opnås et mættet teoretisk system (2), hvor alle tilstande er klassificeret enten kemiske eller ikke-kemiske (fysiske).

SEMANTISKE NORMER

I forrige afsnit blev nogle syntaktiske normer nævnt. Her skal nævnes to normer, der drejer sig om syntaksens forhold til semantikken.

Model

En model for et formelt system er en semantisk struktur, der tolker alle det formelle systems aksiomer som værende gyldige.

Fuldstændighed

Meget ofte er det tilfældet, at en model udover at være bestemt som løsninger til det formelle system også indeholder mere information. Som omtalt viste Gödel, at der i aritmetikkens formelle system altid kunne konstrueres en formel, hvor hverken formelen selv eller dens negation kunne udledes af systemet. Alligevel kunne den vurderes som sand eller falsk i den sædvanlige model for dette aritmetiske system.

Man siger også at aritmetikkens formelle system udover at være umættet (syntaktisk norm) er ufuldstændigt (semantisk norm) over for sin normale model.

Omvendt vil man sige at hvis modellen kun indeholder løsninger til det formelle system er det formelle system fuldstændigt i forhold til modellen.

RESUMÉ

I dette kapitel er forholdet mellem system og model behandlet for meget simple systemer.

Såvel dokumenter som emnedatasystemer kan betragtes som systemer, hvis mening eller tolkning foregår via modeller.

Men dokumenter og emnedatasystemer er særdeles komplicerede systemer, og enhver generalisation fra simple til komplicerede systemer skal gøres med største omtanke og forsigtighed. Derfor er de simple systemer her kun anvendt til begrebsafklaring, mens egenskaberne ved dokumenter og emnedatasystemer er behandlet for sig selv.

Forholdet mellem system og model er behandlet i den formelle logik. Det er et grundlæggende vilkår for de behandlede simple systemer, at systemet er umættet, hvilket medfører

-at der inden for et systems rammer altid kan konstrueres udsagn, der ikke kan udledes af systemet

-at et givet system altid er ufuldstændigt over for sin model, d.v.s. det forklarer ikke hele modellens eller tolkningens indhold.

Vi har set eksempler på, hvorledes emnedatasystemer og dokumenter begge er umættede og ufuldstændige systemer.

Dette illustrerer, at intet dokument eller emnedatasystem udtaler sig udtømmende om sit emne. Specielt er emnedatasystemer blot en listning af ord med få eller ingen relationer imellem sig, og de få "indre sammenhænge", der eksisterer i et sådant system, "smuldrer" efterhånden som systemet tilpasses via ad hoc justeringer til den videnskabelige udvikling.

KONKLUSION

Konsekvensen er at modeller eller tolkninger af såvel dokumenter som emnedatasystemer altid er åbne over for andre systemer end de eksplicit nedskrevne.

Men det er også klart, at jo større den "indre sammenhæng" er i dokumentet eller i emnedatasystemet, jo færre tolkningsmuligheder er der for indekseren.

Et eksempel er stærkt "teori"- eller systemdefinerede konstruktioner som "Naturvidenskab i almindelighed".

3. INFORMATIONSTEORETISKE BETRAGTNINGER

Såvel dokumenter som emnedatasystemer siger vi indeholder information - det er derfor vi foretager informationssøgning.

Hvis det ikke er forkert, er det i alle tilfælde upræcist.

Læserne af dokumenter og emnedatasystemer kan have vidt forskellige perspektiver på deres læsning. Den information, der skulle blive resultatet af læsningen er ikke udelukkende dokumentbåret, som det forhåbentlig er fremgået af de foregående afsnit. Læserens interesser, forhåndsviden og forudsætninger spiller ind på de mulige tolkninger eller modeldannelser.

Traditionel informationsteori beskæftiger sig med at analysere selve meddelelsen, svarende til selve dokumentet. Ved hjælp af nogle få begreber fra den traditionelle informationsteori kan der dannes et mål for et dokumentets eller et emnedatasystems informationsmængde. Senere skal vor - ganske vist sparsomme - viden om selve tolkningsprocessens informationsmængder danne grundlag for en sammenligning mellem selve dokumentet og dets tolkning.

INFORMATIONSINDHOLD PÅ SYSTEMNIVEAU

De begreber, der præsenteres her, siger intet om informationens *mening*; ganske som et teoretisk system intet siger om dets mening, eller om hvad der er sandt eller falsk.

En meddelelses informationsindhold kan måles, men de mål, der dannes, er afhængige af beskrivelsesniveauet. En meddelelse, et dokument eller et emnedatasystem kan eksempelvis beskrives på et binært, alfabetisk, verbalt eller et syntaktisk niveau (se Tabel I).

Niveau	binært	alfabetisk	verbalt	syntaktisk
Enhed	bits	bogstaver	ord	sætninger

TABEL I

En meddelelses informationsindhold på forskellige niveauer med tilhørende måleenhed.

I den binære beskrivelsesform måles informationen som det antal pladser (cifre) i et to-tals system, der skal anvendes for at skrive meddelelsen.

EKSEMPEL

Skrives informationen 7842 i vort sædvanlige ti-talsystem, ser det i et to-talsystem ud som

$$1111010100010 = 1 \times 2^{12} + 1 \times 2^{11} + \dots + 1 \times 2^1 + 0 \times 2^0$$

Denne information fylder 13 bits, svarende til at den giver svar på 13 spørgsmål af typen: Er det første ciffer 0? Er det andet ciffer 0?...Alle spørgsmålene kan besvares med ja eller nej.

Intuitivt er det indlysende, at jo flere spørgsmål, det er muligt at stille og få svar på i forhold til en given meddelelse, jo mere information må meddelelsen indeholde.

Vedtager vi, at det er denne informationskapacitet, der er vort mål for information, har vi dannet et additivt informationsmål.

Vender vi nu tilbage til tallet 7842 kunne det skrives på en mere bekvem måde, hvis vi udnytter det normale positionstalsystem. De enkelt cifre skrives da som 7=111, 8=1000, 4=100 og 2=10. Sættes disse tal uden videre sammen til én meddelelse lyder den som 111100010010. Men nu ved vi ikke om det første ciffer er 1=1, 11=3 eller 111=7. Derfor må der indføres den konvention, at hvert ciffer fylder det maksimale antal bits, der er nødvendige for at beskrive tallene 0 til 9, og det er 4. Den resulterende meddelelse, der før fyldte 13 bits fylder nu 16 bits! Til gengæld er det lettere at læse.

"Overflødige bits" kaldes redundans. Redundans gør teksten lettere at læse, dvs. det kræver mindre forhåndsviden at læse teksten. Redundans gør også teksten mere "robust" overfor støj.

På samme måde som de almindelige tal kan oversættes til det binære talsystem, kan alfabetet oversættes hertil.

På dansk har vi 29 bogstaver samt mellemrum, punktum og komma - i alt 32 tegn. Disse tegn kan hver for sig udtrykkes ved 5 binære cifre eller 5 bits.

I praksis er der hyppige og sjældne bogstaver, hvorfor det er muligt at lave en kode, der forbinder de oftest anvendte bogstaver med den korteste kode og til gengæld tildeler de sjældnest anvendte bogstaver den længste kode. Hermed kan man nedbringe det gennemsnitlige forbrug af bits pr. bogstav til 4,8. Tages der derudover hensyn til, at man ikke behøver at læse alle bogstaver for at forstå en meddelelse - meddelelsen *sandstrnd* forstås til trods for, at der mangler et bogstav - er det gennemsnitlige informationsindhold ca. 2 bits/bogstav, mens mere systematiske sprog som tysk er endnu længere nede.

BÅNDBREDDE FOR LÆSNING

Med et informationsindhold på 2 bits pr. bogstav bliver den rent maskinelle læsehastighed, beregnet som én normalside á 2400 tegn på 2 minutter, til 40 bits pr. sekund.

Men læsning består ikke blot i at læse dokumenternes tekst. Teksten skal også fortolkes.

BÅNDBREDDE FOR TOLKNING AF DOKUMENTER OG EMNEDATASYSTEMER

Hvad sker der egentlig i informationsteoretisk forstand på tolkningsniveauet? Hvad sker der når vi bevæger os væk fra de trivielle systembundne opregninger af informationsindhold til meningen med informationen?

Kan det beskrives på tilsvarende måde?^a

^a I informationsteorien forsøges dette. Det giver store problemer idet tolkningsniveauernes informationsindhold er mindre end de "maskinelle" systemniveauer, idet redundansen falder op gennem niveauerne, udtrykt ved at informaa (fortsættes...)

Nej - netop ikke! Tolkning eller skabelse af en model af et dokument eller emnedatasystem afhænger af mange andre faktorer. Det afhænger af læserens samlede viden, historie og sprog m.v..

En præcis beskrivelse af sådanne tolkninger informationsindhold kan vi ikke gå ind på. Men det er muligt at argumentere for sådanne tolkninger båndbredder. Det er muligt at argumentere for hvor informationsrige de processer er, der danner tolkningerne og modellerne, hvorved vi får et indtryk af tolkningernes kompleksitet. Informationsbehandlingen i menneskets hjerne antages sædvanligvis at involvere mellem 1 og 100 sekventielle trin pr. sekund^a. Dette er en lav hastighed i sammenligning med at en hurtig PC's clock-frekvens nærmer sig 100 MHz.

Alligevel ved vi, at mennesker på mange områder er selv den største computer langt overlegen - og formentlig altid vil være det. Det skyldes blandt andet at den menneskelige hjerne arbejder med mange samtidige beregninger. Det kaldes parallel informationsbehandling.

Antallet af samtidige beregninger skønnes konventionelt at være omkring 10^{12} trin/sekund, svarende til en informationsbehandlingskapacitet på - lavt sat - 10^{12} bits/sekund. Dette skøn kan være 100 gange for stort eller 100 gange for lille. Alligevel viser det den enorme informationsbehandlingskapacitet, den menneskelige hjerne er udstyret med - en informationsbehandlingskapacitet, der er til rådighed og karakteriserer mulighederne, når mennesker tænker, får idéer og foretager fortolkninger.

Sammenlignes tolkningskapaciteterne på 10^{12} bits pr. sekund med læsekapaciteten på omkring 40 bits pr. sekund, ser vi en næsten svimlende overkapacitet i menneskets muligheder for at tolke og få idéer.

^a(...fortsat)

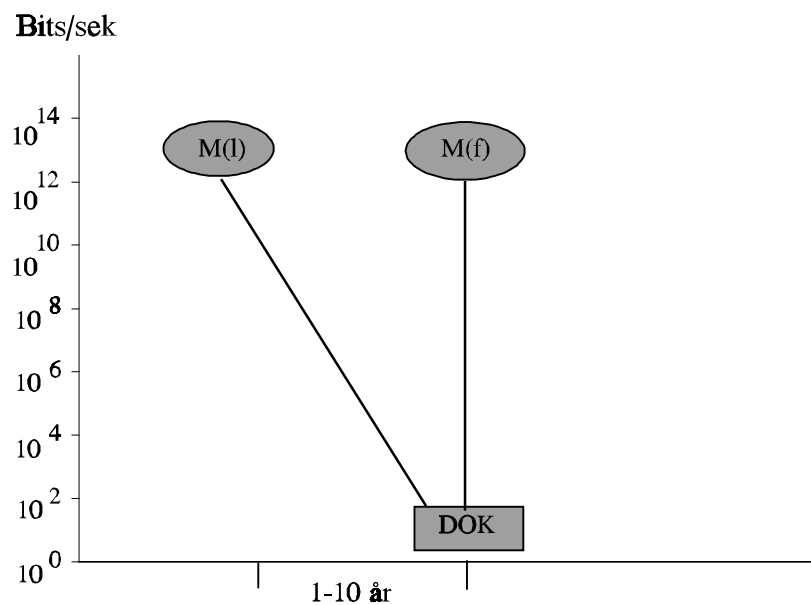
tionsindholdet antager sin maksimalværdi, når vi intet ved på forhånd om svarenes sandsynlighed, d.v.s. når svarene synes tilfældigt fordelt.

^a Bernsen & Ulbæk 1991:211 ff., Brunak & Lautrup 1988:32 ff. eller Rumelhart, McClelland & PDP 1989:part 1:130 ff...

Bredes disse iagttagelser ud til at dække emnesøgningsproblematikken, ser vi en kompliceret række af spring mellem den menneskelige hjernes meget store kapacitet og skriftens enkelthed.

AT LÆSE ET DOKUMENT

Tag den simple situation, hvor en person læser et dokument. Dette kan betragtes som en information fra forfatteren til læseren via det skrevne dokument, som anskueliggjort i Figur 6.



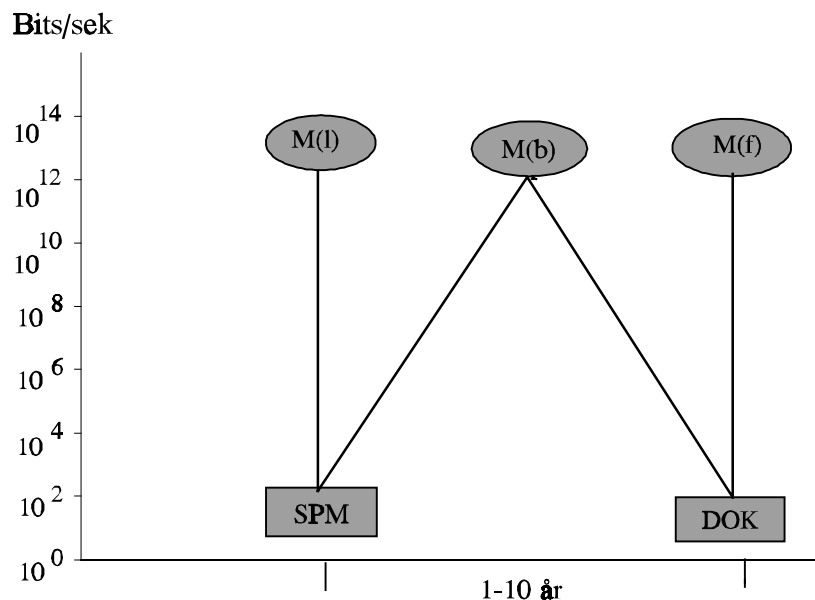
FIGUR 6 Beskrivelse af spring i informationskapaciteter ved læsning af et dokument

Forfatterens tolkning, $M(f)$, og læserens tolkning, $M(l)$, af værket, DOK, foretages på et en meget høj båndbredde på 10^{12} - 10^{16} bits pr. sekund, mens selve dokumentet, der er kommunikationen mellem forfatterens og læserens tolkning, foregår på læsningens båndbredde på ca. 40 bits pr. sekund.

Umiddelbart synes det som en umulig opgave, at læseren skulle kunne rekonstruere forfatterens mening med dokumentet - simpelt hen på grund af niveauforskellen i informationskapacitet. På den anden side ved vi, at det kan lade sig gøre, men det kræver, at der etableres et vist fælles mål af tolkningsrammer.

Læseren har en stor sandsynlighed for at kende forfatterens perspektiv, hvis de tilhører samme kulturkreds og er samtidige. Såfremt der er tale om andre kulturkredse eller ældre tiders forfattere, vil det ofte kræve ekstra arbejde af læseren, at sætte sig ind i forfatterens perspektiv.

AT FINDE ET DOKUMENT MED EN BIBLIOTEKAR SOM MELLEMLED
Antag, at der er tale om en forespørgsel til en bibliotekar. Da indskydes et led mellem læseren og dokumentet - et led, der skal fortolke såvel bogens emne som læserens behov, som det er kommet til udtryk via et spørgsmål. Denne proces er anskueliggjort i Figur 7.



FIGUR 7 Beskrivelse af spring i informationskapaciteter ved en læsers forespørgsel til en bibliotekar for at finde et dokument om et emne.

Læserens spørgsmål til bibliotekaren, SPM, dækker over en mening med dette spørgsmål, $M(l)$, der er dannet med høj informationsbe-
arbejdningskapacitet. Dette spørgsmål tolkes af bibliotekaren på
samme høje kapacitetsniveau, $M(b)$. Bibliotekaren har også en
tolkning, $M(b)$, af dokumentet, DOK, dannet med høj kapacitet,

ganske som det er tilfældet for forfatterens tolkning, $M(f)$, af sit værk.

Allerede her er der mange muligheder for at det enkle spørgsmål, der ofte er udtryk for en kompliceret problemstilling hos læseren, af bibliotekaren tolkes på anden måde end læseren mente. Med denne - korrekte eller ukorrekte - tolkning skal bibliotekaren nu prøve at finde et værk, der, i bibliotekarens tolkning af forfatterens intentioner, stemmer overens med bibliotekarens opfattelse af læserens spørgsmål.

Her - som før - er den centrale forudsætning for at formidlingen lykkes, at forfatter, læser og bibliotekar tilhører samme kultur og samme tid, eller at læser og bibliotekar arbejder på ikke blot at danne sig et billede af forfatteren, men også af hinanden.

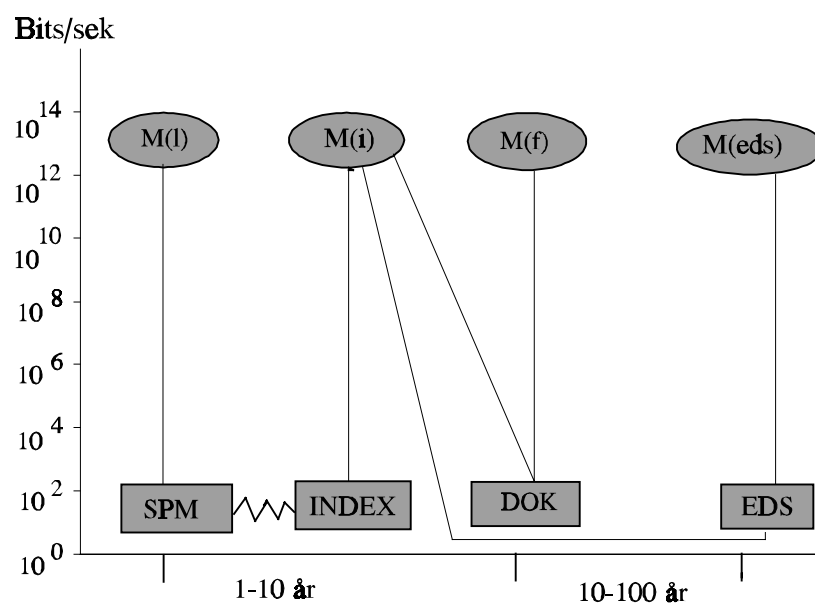
Det lykkes overraskende tit, fordi mennesker indlever sig utrolig hurtigt, når de vel at mærke har noget at forholde sig til.

AT FINDE ET DOKUMENT MED ET EMNEDATASYSTEM SOM MELLEMLIED

Her indføres en emneregistrering af dokumentet som et emnebeskrivende led mellem forfatter og læser. En emneregistrering består af to processer og en kompliceret samkøring af disse to processer. Dels skal dokumentet tolkes af en indeksør og dels skal emnedatasystemet tolkes af indekseren. Til sidst skal indekseren kombinere disse to fortolkninger ved at udvælge én eller flere koder fra emnedatasystemet som fortolkning af dokumentets emne.

Dette er vist på Figur 8.

Læseren stiller et spørgsmål, SPM, der dækker over en mening, $M(l)$, med dette spørgsmål, dannet med høj informationsbehandlingskapacitet. Dette spørgsmål udvælger i emnesøgningsprocessen en emneindeksering, INDEKS. Denne emneindeksering er typisk foretaget 1-10 år før af en indeksør, der har haft en tolkning, $M(i)$, af den foretagne emneindeksering, foretaget med høj informationsbehandlingskapacitet. Med samme kapacitet har indekseren tolket dokumentet, DOK, og emnedatasystemet, EDS. Såvel dokumentet som emnedatasystemet er dannet ud fra et perspektiv og en tolkning hos de respektive forfattere, foretaget med høj informationsbehandlingskapacitet.



FIGUR 8 Beskrivelse af spring i informationskapaciteter i en læsers spørgsmål, der via en indekssørs emneindeksering af et dokument med et kontrolleret emnedatasystem søger at formidle forfatterens tolkning af sit dokument emne.

Hvis processen før var kompliceret gives der intet ord for denne ganske normale formidlingspraksis via bibliotekernes online kataloger.

Ikke blot skal der dannes et fælles perspektiv mellem forfatter, indekssør og læser, men også emnedatasystemets tolkning indgår i denne fælles referenceramme. Det lader sig undertiden gøre, til trods for at læseren ikke har nogen mulighed for at få kendskab til indekssørens referenceramme og sjældent sætter sig ind i emnedatasystemets fortolkningsramme. Det lader sig gøre, hvor der er tale om et fælles centralperspektiv. Men det er ikke overraskende, at der i denne yderst sammensatte kommunikationsproces optræder en god del støj.

KONSEKVENSER FOR EMNE- INDEKSERING I BIBLIOTEKSKATALOGER

I det følgende sammenfattes konsekvenserne af de tre første kapitlers empiriske undersøgelser og overvejelser over emnesøgning og emneindeksering i bibliotekskataloger.

Undersøgelserne af bibliotekskataloger viste, at brugerne har store vanskeligheder med at foretage emnesøgninger i bibliotekernes online kataloger, og anvender dem mindre og mindre til dette formål. Vi så også, at der var tale om et indholdsproblem - ikke et præsentationsproblem. Emneindekseringen skulle være betydelig mere informativ i form af indholdsfortegnelser, abstracts m.v.. Men dette skaber et andet problem, nemlig at den større tekstmængde i litteraturbeskrivelsen resulterer i at brugernes i forvejen for store søgesæt vokser yderligere.

Den forbedrede emneindeksering skaber altså blot - mindst - ét andet problem. Dette kunne løses ved at reducere antallet af poster, der stilles til rådighed for emnesøgning - altså et brud på den hidtidige konvention, at alle bøgerne i bibliotekskataloger emneindekseres.

Men den forbedrede emneindeksering skaber også et andet problem: Ved at lade indholdsfortegnelser og abstracts udgøre dokumenters emnebeskrivelse, foretages emnebeskrivelse og emnesøgning i naturlig tekst i stedet for de traditionelle emnebeskrivelser i kontrollerede sprog.

Kontrollerede sprog tilbyder generelle eller syntetiserende begreber og entydige emnebeskrivelser. Men disse fordele hviler på, at der

indekseres konsistent - at der vælges de "rigtige" beskrivende koder eller ord i det kontrollerede sprog. Gennem talrige undersøgelser er det vist, at denne konsistens ikke eksisterer i praksis, og at man heller ikke ved hvordan den kunne øges, samtidig med at det er højst usikkert, om konsistent emnebeskrivelse med kontrollerede sprog har nogen som helst indflydelse på genfindingsprocessen.

Essensen af i de videnskabsteoretiske og informations-teoretiske undersøgelser af emnebegrebet og emneindeksering er, at der kræves en helt anderledes autentisk emnebeskrivelse, end vi er vant til at nøjes med i bibliotekskataloger.

Det er grundigt illustreret, at der ikke gives muligheder for éntydige emnebeskrivelser og éntydige tolkninger af dokumenter eller emnedatasystemer^a. Ingen systemer - og slet ikke så åbne eller umættede systemer som dokumenter og emnedatasystemer - er så udviklede eller mættede, at de kun giver plads til én tolkning.

I mangel af middelalderens centralperspektiv, er betingelsen for at brugeren kan rekonstruere dokumenternes emner i online kataloger, at brugeren kan få viden om det perspektiv, dokumentbeskrivelsen er foretaget ud fra.

Dette kan gøres ved at anvende oversigtslitteraturens organisering af dokumenter i emner som emneindeksering.

Den samlede konklusion er, at forbedringer af emnesøgningsmulighederne i bibliotekskatalogerne går gennem en dobbelt strategi. Denne dobbelte strategi består i at lade emnesøgning foregå i den delmængde af den anskaffede litteratur, der bredt kunne

^a Der er mange måder, at illustrere dette på. Én af de mere uheldige er, at parallelisere med Heisenbergs usikkerhedsrelationer, som det er gjort flere steder i den informationsteoretiske litteratur, selv af så velovervejede forfattere som Marcia Bates (Bates 1986). Det er populært at anvende naturvidenskabelige analogier, men i dette tilfælde har analogien intet med det, den skal illustrere, at gøre. I informationsteorien er der ikke, som i fysikken, nogen principiel udelelig enhed som Planck's konstant, der kan retfærdiggøre analogien. Og hvorfor gå over åen efter vand, når inkonsistent emneindeksering skyldes, at det ikke er muligt at lave en fuldstændig beskrivelse af systemets tilstand. Dette er en klassisk problematik. Den klassiske fysik kan ikke foretage en fuldstændig beskrivelse af et systems "begyndelsesbetingelse". Men ironisk nok er dette netop muligt i kvantemekanikken. Så analogien er direkte misvisende.

kaldes oversigtslitteraturen, og foretage emneregistreringen ved at genbruge dele af dokumentets tekst så som indholdsfortegnelse, abstracts eller stikordsregister.

Det er et alvorligt brud med den hidtidige praksis i emneindekseringen af online kataloger.

I resten af dette arbejde vil vi beskæftige os med emneindeksering ud fra ovenstående principper. Først vil analytiske kataloger blive gennemgået, og endelig vil den emneindekseringsform, der er kaldet PARADOKS-indeksering, blive beskrevet og evalueret.

ANDEN DEL

PRAKTISKE ANVENDELSER

ANVENDELSE AF NATURLIG TEKST TIL EMNEBESKRIVELSE I ONLINE KATALOGER

Dette kapitel handler om anvendelser af - i det væsentlige - indholdsfortegnelser som beskrivelse af dokumenters emne i bibliotekskataloger.

Behandlingen vil specielt rette sig mod anvendelse af indholdsfortegnelser til at beskrive indholdet i samleværker - et træk, der berettiger til navnet "analytiske kataloger".

Indholdsfortegnelser som emnebeskrivelse i bibliotekskataloger er langt fra nyt.

Man har altid vidst, at et godt katalog indeholdt indholdsfortegnelser for, på simpel vis at fortælle brugeren, hvad bogen indeholder - ikke ved hjælp af fremmedartede systematiske koder og emneord, men med forfatterens egne ord.

Problemet har været teknologisk. Indholdsfortegnelser i bibliotekskataloger var af begrænset værdi før indførelse af edb-teknologi.

I kortkatalogets dage blev der - mere eller mindre sporadisk - foretaget analyser af bøgerne. Hele indholdsfortegnelsen - eller dele heraf - blev opført på kortet. Men der var meget få opslagsmuligheder til denne berigelse; undertiden blev der lavet nogle få systematiske biplaceringer eller tilføjet nogle emneord som via ekstra kort kunne lede til analysen. Det var ikke teknisk muligt, at give muligheder for opslag på indholdsfortegnelsernes ord - specielt ikke hvis ordene stod inde i artiklens titel. Årsagen er, at hver ny opslagsmulighed kræver ét ekstra kort.

Det var - i det væsentlige - kortkatalogets teknologiske muligheder for at give emnesøgningsmuligheder på samleværkers indhold^a.

En moderne anvendelse af indholdsfortegnelser som emnebeskrivelser i bibliotekskataloger er repræsenteret i den såkaldte SAP-indeksering.

SAP-INDEKSERING

SAP-indeksering er en emneindekseringsmetode, der anvender betydende ord, udvalgt fra indholdsfortegnelser og stikordsregistre som emnebeskrivelse efter bestemte udvalgskriterier.

SAP-indeksering blev udviklet og delvist afprøvet af Pauline Atherton i projektet med navnet Subject Access Project^b.

I det oprindelige projekt blev denne udvælgelse gjort mere eller mindre automatisk og reproducerbar ved hjælp af kvantitative udvælgelseskriterier - f. eks. blev det krævet, at ordene skulle henvise til et vist minimum af sider for at blive medtaget.

Forsøget blev foretaget på en bogmasse på ca. 2400 titler.

Bøgerne udvalgte ligeligt indenfor humaniora og samfundsvidenskab blandt bøgerne på et par videnskabelige biblioteker ud fra den systematiske placering, de allerede havde fået.

Senere blev SAP-indekseringen videreudviklet af Irene Wormell^c. Emnebeskrivelsen blev i dette projekt ikke blot begrænset til udvalg blandt indholdsfortegnelse og stikordsregistrets ord, men indbefattede også figur- og tabeltekster. De kvantitative udvælgelseskriterier blev også gjort mindre rigide. Forsøget omfattede 1.100 svenske ministerielle udvalgsrapporter, publiceret mellem 1972 og 1981. Fra 1983 er denne indekseringsmetode anvendt i denne offentligt tilgængelige database "RÄTTSDATA", der indeholder ministerielle dokumenter.

^a En detaljeret historisk gennemgang findes i Hyman 1978.

^b Atherton 1977, Cochrane 1985.

^c Wormell 1985.

På Purdue Universitets bibliotek er monografi-posterne for mere end 20.000 monografier indenfor ingeniørvidenskab berigede med udvalgte ord fra indholdsfortegnelser, og det rapporteres, at den mere specifikke og opdaterede terminologi fra indholdsfortegnelserne, er de traditionelle klassifikationssystemer overlegne^a.

Alex Byrne og Mary Mico har lavet en SAP-baseret registrering for det australske forsvarsakademi's bibliotek, der både viser styrker og svagheder ved disse registreringsformer^b.

SAP-indeksering er indeksering med systematisk klassifikation og emneord overlegen på følgende områder:

- Emnesøgningerne resulterede i flere relevante titler mens søgepræcisionen var den samme^c.
- Det gik dobbelt så hurtigt at foretage emnesøgningerne^d.

SAP-registreringens problemer er:

- Flere relevante titler kan også - som det er påpeget tidligere^e - være et stort problem, idet fundene bliver for store til, at det er muligt at foretage en udvælgelse ved simpel gennembladrning. Dette understreges af at Byrne & Micco fandt at SAP-registreringen øgede genfindingen med 300%!^f.
- Dårlige indholdsfortegnelser og stikordsregistre. Således registrerede Atherton at omkring 10% af dokumenterne enten

^a Posey & Erdmann 1986.

^b Byrne 1986, Byrne & Micco 1988.

^c Atherton 1977 og Cochrane 1985.

^d Atherton 1977 og Cochrane 1985.

^e Se side 19.

^f Byrne & Micco 1988.

manglede eller havde uegnede indholdsfortegnelser og stikordsregistre^a.

- De kvantitative udvælgelseskriterier for udvælgelse af ord og sætninger i indholdsfortegnelser og stikordsregistre blev vurderet som værende uhensigtsmæssige^b.

I det følgende skal vi udelukkende beskæftige os med analytiske kataloger.

1. ANALYTISKE KATALOGER

Bibliotekskataloger i Danmark registrerer bind - den enhed, der anskaffes, systematiseres, katalogiseres, udlånes, hjemkaldes, reserveres, indbindes, repareres og kasseres. En praktisk enhed for en række processer i et bibliotek.

Brugeren er interesseret i værker: en titel af en forfatter, der skriver om et emne.

I tidsskrifter samles en række værker altid.

I bøger er der normalt ét værk i hvert bind. Men undertiden er der flere værker i en bog. Og så kalder vi det et samleværk.

Samleværker udgives normalt om et bestemt emne eller bestemte emnekredse. Der er en fagligt ansvarlig redaktør, som har stået for samling og redaktionel bearbejdning af bidragene. Det har den fordel, at der er samlet en række forskellige beskrivelser af tilsyneladende samme emnekreds. Emnet beskrives udover af redaktøren også gennem de forskellige bidrag. Det gør samleværker specielt værdifulde i forbindelse med emnesøgninger.

Samleværker behandles normalt som ét værk i et online katalog. De enkelte værker eller bidrag kan ikke findes i online kataloget, til

^a Atherton 1977 og Cochrane 1985.

^b Atherton 1977 og Cochrane 1985.

trods for at brugeren normalt er interesseret i værker ikke i bind, til trods for at samleværkets samling af enkelte bidrag via disses forskellighed tilbyder en fyldig beskrivelse af emnet, og til trods for at det nogen gange er lidt af en tilfældighed, om de enkelte bidrag i et samleværk er udgivet samlet eller separat.

Bibliotekets online katalog er praktisk for de interne processer, men ikke praktisk for brugeren.

Til trods for at der er udvist en betydelig interesse for berigelse af poster med indholdsfortegnelser både fra brugeres side og biblioteksvidenskabelig side, er man tilbageholdende med at realisere denne udvidelse af katalogerne.

Hindringen er tilsyneladende mangel på resurser eller - rettere sagt resurseprioritering.

Hovedargumentet mod en berigelse af vore online kataloger er, at der allerede findes fyldigere litteraturbeskrivelser for de fleste emners vedkommende i den bibliografiske litteratur, herunder de flittigt brugte online tilgængelige tidsskriftbibliografier. Det er for dyrt at opfylde brugernes behov i det lokalt producerede online katalog, når der eksisterer centralt producerede faciliteter til emnesøgning.

En i mange henseender resurse-økonomisk løsning. Men også en løsning, der nok en gang prioriterer tidsskrifter højere end bøger med det resultat, at bøgerne, der trods alt optager en del af bibliotekernes budgetter, bruges mindre end det ville være tilfældet, hvis de var lige så godt beskrevet som tidsskrifterne er.

Konsekvensen er, at brugerne er henvist til at foretage de mere krævende emnesøgninger i trykte og edb-baserede bibliografier, mens de helt påviseligt ønsker, at de eksisterende online kataloger beriges med fyldigere litteraturbeskrivelser, så de kan benyttes til emnesøgning i de forhåndenværende samlinger - også bøgerne.

Hvis bibliotekerne vil tage brugernes behov alvorligt skal online katalogerne beriges.

Det skaber behov for en ændret resurse anvendelse.

Og hvis man overvejer at berige eksisterende online kataloger, er det vigtigt, at se på bibliotekernes nuværende anvendelse af resurser.

RESURSER TIL AT HJEMSKAFFE OG TILGÆNGELIGGØRE BØGER

Produktion af online kataloger er det dyreste led i hele emnesøgningssystemet.

Denne produktion kan groft taget deles op i tre dele: anskaffelse, grundkatalogisering og emnebeskrivelse.

ANSKAFFELSE af litteratur, herunder udvælgelse og indkøb, tager i gennemsnit 1 mandetime pr. bog^a. Her og i det følgende er udtrykket "mandetimer" anvendt som den totale arbejdstid, inklusive sygdom, ferier, pauser m.v. - altså et udtryk for den krævede bemanning. Så selv om et arbejde tager ½ time at udføre, vil det kræve ca. 1 mandetime.

GRUNDKATALOGISERING koster et par hundrede kroner hos Dansk Bibliotekstjeneste, hvilket er af samme størrelsesorden, som det skønnes af Library of Congress^b og Mats Cavallin^c.

Nu genbruger bibliotekerne ofte hinandens poster i on-line katalogerne. Det reducerer selvfølgelig den gennemsnitlige pris for at producere en on-line post.

Med en gennemsnitlig genbrugsprocent på lidt over 50% regner Det Kongelige Bibliotek med en gennemsnitlig tid til grundkatalogisering på 1 mandetime^d.

EMNE- OG INDHOLDSBESKRIVELSEN af bogen - udvikling og formidling af det systematiske beskrivelsessystem og et eventuelt emneordssystem beløber sig til mindst 90 kr. pr. bog hos Library

^a Det Kongelige Bibliotek 1989.

^b Mandel 1985.

^c Cavallin 1991:22.

^d Det Kongelige Bibliotek 1989.

of Congress, beregnet i 1985-priser^a. Her er en af de vægtigste omkostninger, vedligeholdelse og opdatering af den faglige specialists viden ikke medregnet. Dette er i overensstemmelse med at det gennemsnitlig tager 1 mandetime at systematisere og færdiggøre bogen, inklusive opdatering af systemer og vedligeholdelse af den faglige specialists viden^b.

Med den eksisterende praksis på bibliotekerne betyder det således, at de samlede behandlingsomkostninger for en bog - lavt sat - er 3 mandetimer. Det svarer til at et middelstort forskningsbibliotek med en bogaccession på 8.000 bind om året anvender 24.000 mandetimer eller omkring 3 mill. kr. om året til at gøre bøgerne genfindelige.

ÆNDRET RESURSEANVENDELSE

Der anvendes således mere end det dobbelte af bogbudgettet til at tilgængeliggøre bøgerne. En så stor omkostning opmuntrer ikke til at finde veje til at øge samme post på budgettet ved at udbygge emnebeskrivelsen.

Men en nærmere undersøgelse viser, at en ændret resurseanvendelse ville kunne give plads til at lægge indholdsfortegnelser af samleværker i bibliotekernes online kataloger.

HVOR MANGE VÆRKER SKJULER ET BIBLIOTEKSKATALOG?

Brugeren ser i bibliotekets online katalog færre titler end biblioteket egentlig har.

For at kunne vurdere konsekvenserne af, at de enkelte bidrag i samleværkerne beskrives i online katalogerne, må vi kende antallet af samleværker samt det gennemsnitlige antal bidrag i disse samleværker.

^a Mandel 1985.

^b Det Kongelige Bibliotek 1989.

Der er foretaget undersøgelser af dette spørgsmål på Santa Ana College Library i Californien^a, som senere inspirerede til tilsvarende undersøgelser på Danmarks Pædagogiske Bibliotek^b og Roskilde Universitetsbibliotek^c. Disse undersøgelser skal refereres her, da de giver helt centrale oplysninger om størrelsen og omkostningerne ved denne type forbedringer af bibliotekskatalogernes emnesøgningsfaciliteter.

SANTA ANA COLLEGE LIBRARY

Dette bibliotek er et universal- og institutionsbibliotek for en højere læreanstalt med 70.000 bind.

En stikprøve på 4.094 bøger viste at 871 var samleværker, svarende til 21.3%. Samleværkerne indeholdt i gennemsnit 31.2 titler.

For hver titel i ovenstående bibliotekskatalog gemmer der sig således 6 til i bogmassen. Biblioteket kan tilbyde 600% flere titler at søge i, end i dag, uden at købe én bog mere. Eller: lagt op i et online katalog svarer det til, at der i stedet for 70.000 er omkring 350.000 søgbare titler, stadig uden at købe flere titler.

DANMARKS PÆDAGOGISKE BIBLIOTEK

Danmarks Pædagogiske Bibliotek (DPB) har en afdeling - nu kaldet Pædagogisk Specialsamling - der er et højt specialiseret forskningsbibliotek med ca. 350.000 monografier indenfor pædagogik og børne- og ungdomspsykologi. Den årlige tilvækst er på ca. 8.000 bøger.

Der blev udtaget stikprøver på 3 forskellige måder:

STIKPRØVE 1: ALLE BØGER PUBLICERET FØR 1990.

På DPB er alle bøger, publiceret før 1990 opstillet for sig, da der i 1990 indførtes katalogbrud.

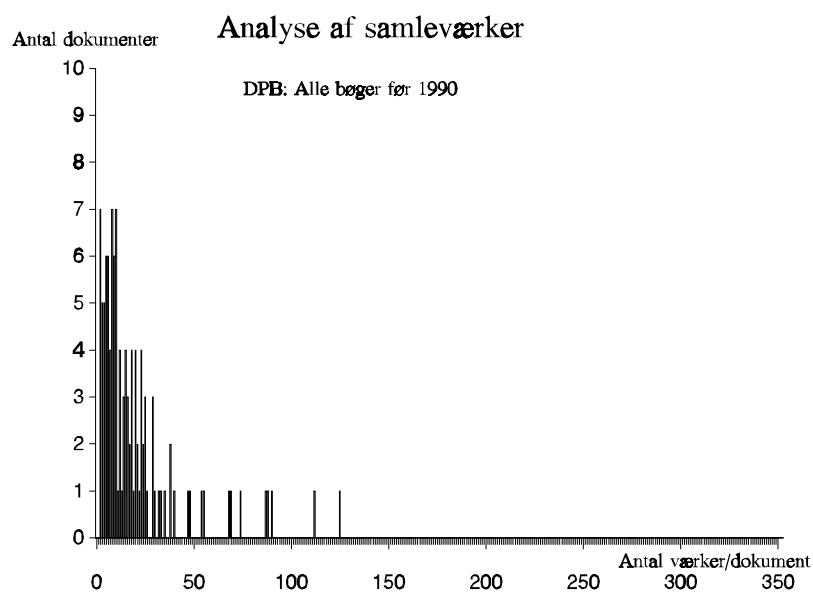
Vi valgte sidste bog fra 4' og 5' hylde på hver reol og udtog herved 718 bind. Antallet af værker blev noteret for hvert bind.

^a Hoffman & Magner 1985.

^b Poulsen & Lenschau-Teglers 1992.

^c Bredsdorff & Poulsen 1993.

Prøven indeholdt 114 samleværker - dvs. samleværkerne udgjorde 16% af prøvens bind. På dette grundlag kan vi med 95% procent sikkerhed sige at der er $16 \pm 3\%$ samleværker i hele samlingen. Samleværkernes bidrag er vist på figur Figur 9. Vi kan på dette grundlag sige, at det gennemsnitlige antal værker pr. samleværk i hele samlingen er 20 ± 4 med 95% sikkerhed^a.



FIGUR 9
Stikprøveanalyse af andelen af samleværker samt antallet af værker i hvert samleværk. Antal prøver: 718. Antal samleværker: 114. (Poulsen & Lenschau-Teglers 1992)

STIKPRØVE 2: ALLE BØGER PUBLICERET EFTER 1990.

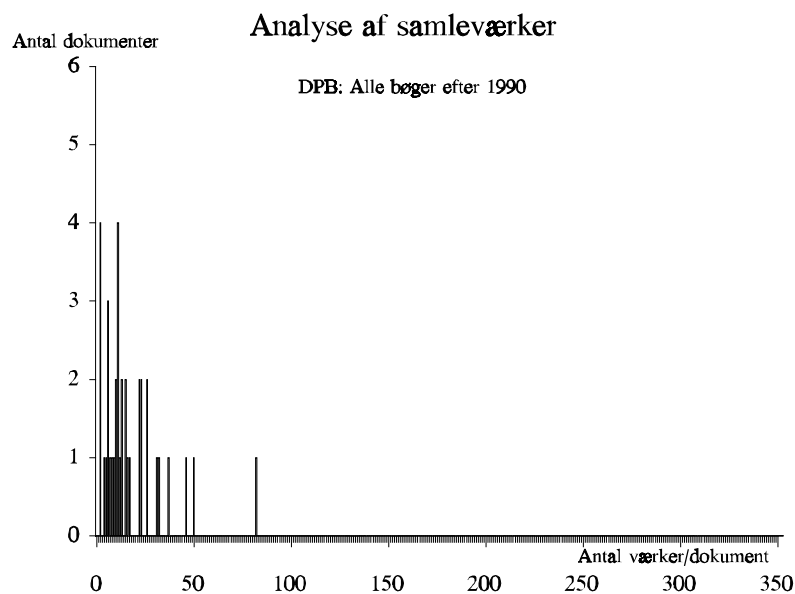
Man kunne forestille sig, at der har været en udvikling af publikations- eller indkøbspraksis. Dette ville da vise sig ved en stikprøveundersøgelse i litteraturen efter 1990 til primo 1992.

I denne periode er 10.000 bøger på Danmarks Pædagogiske Bibliotek færdigbehandlet.

^a Se Appendiks 2.

Vi udtog første og sidste bind på hver hylde. Dette gav en stikprøve på 211, hvoraf 17% var samleværker. På dette grundlag kan vi med 95% sikkerhed forudsige at $17 \pm 5\%$ er samleværker af denne bogmasse.

Antallet af værker i samleværkerne er angivet på Figur 10 . Det gennemsnitlige antal værker pr. samleværk er 17 ± 5 værker indenfor 95% grænsen.



FIGUR 10
Stikprøveanalyse af andelen af samleværker samt antallet af værker i hvert samleværk. Antal prøver: 211. Antal samleværker: 36. (Poulsen & Lenschau-Teglers 1992)

STIKPRØVE 3: ENGELSKSPROGEDE BØGER PUBLICERET 1980-1989

Da mange kongresrapporter er på engelsk og disse rapporter har et betydeligt antal titler, udvalgte vi engelsksprogede bøger publiceret i en ti-års periode for at se, om dette gav anledning til signifikante variationer.

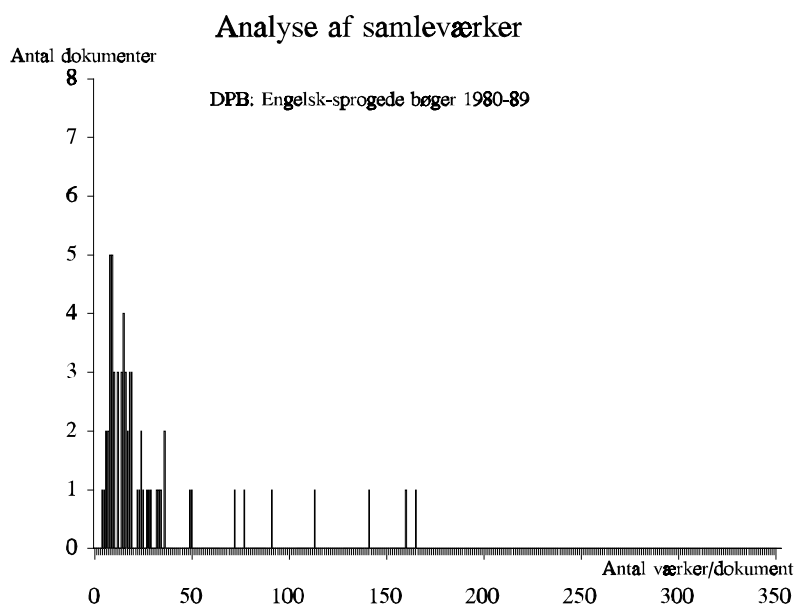
Dette udvalg blev foretaget maskinelt. Det drejede sig om ca. 16.000 bøger.

Stikprøven blev udtaget ved at udvælge poster med to tilfældigt valgte to-cifrede endetal i ISBN nummeret.

Stikprøven viste, at 83 ud af 496 dokumenter var samleværker, hvilket giver 17% samleværker. 8 af disse var udlånt, da stikprøveundersøgelsen foregik, og der kunne ikke foretages en yderligere analyse af disse værker. Vi har derfor ikke medregnet disse, da forskellen er lille sammenlignet med den statistiske usikkerhed.

Regner vi med 75 samleværker, kan vi forudsige at $15 \pm 3\%$ af engelsksprogede publikationer publiceret i 1980-89 med 95% sikkerhed er samleværker.

Samleværkernes bidrag til det totale antal værker er gengivet på Figur 11. Det gennemsnitlige antal værker pr. samleværk var 29 og vi kan med 95% sikkerhed sige, at ovennævnte del af samlingen indeholder gennemsnitlig 29 ± 9 værker pr. samleværk.



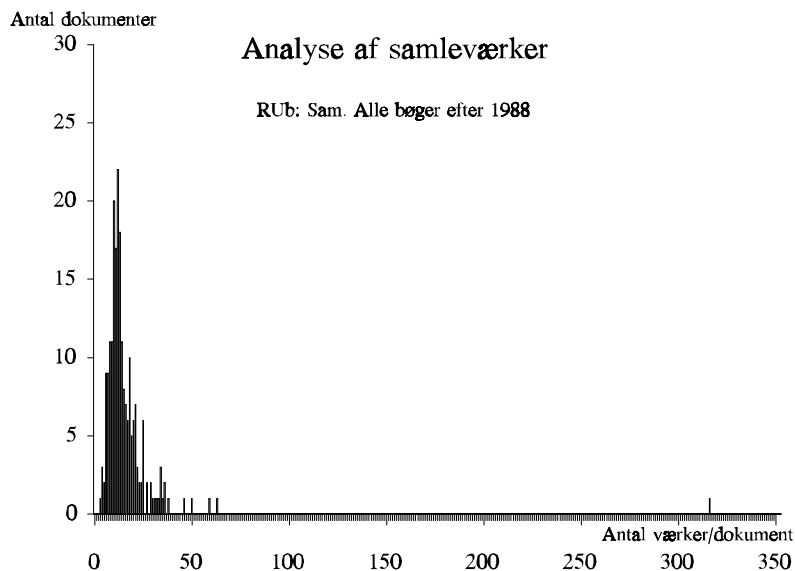
FIGUR 11

Stikprøveanalyse af andelen af samleværker samt antallet af værker i hvert samleværk. Antal prøver: 496. Antal samleværker: 75. (Poulsen & Lenschau-Teglers 1992)

ROSKILDE UNIVERSITETSBIBLIOTEK

Roskilde Universitetsbibliotek (Rub) er et universalbibliotek med ca. 225.000 monografier og en årlig tilvækst på ca. 12.000 monografier.

For at se, om de foregående resultater for DPB er generaliserbare til andre biblioteker, blev der i 1993 foretaget en undersøgelse på dette bibliotek, der adskiller sig markant fra DPB ved at være et universalbibliotek. For se om der skulle være signifikante fagspecifikke karakteristika for stikprøverne blev DPB's hovedsagelig humanistiske fagområder suppleret med de samfundsvidenskabelige og de naturvidenskabelige fagområder på RUB^a.



FIGUR 12

Stikprøveanalyse af andelen af samleværker samt antallet af værker i hvert samleværk. Antal prøver: 887. Antal samleværker: 210. (Bredsdorff & Poulsen 1993)

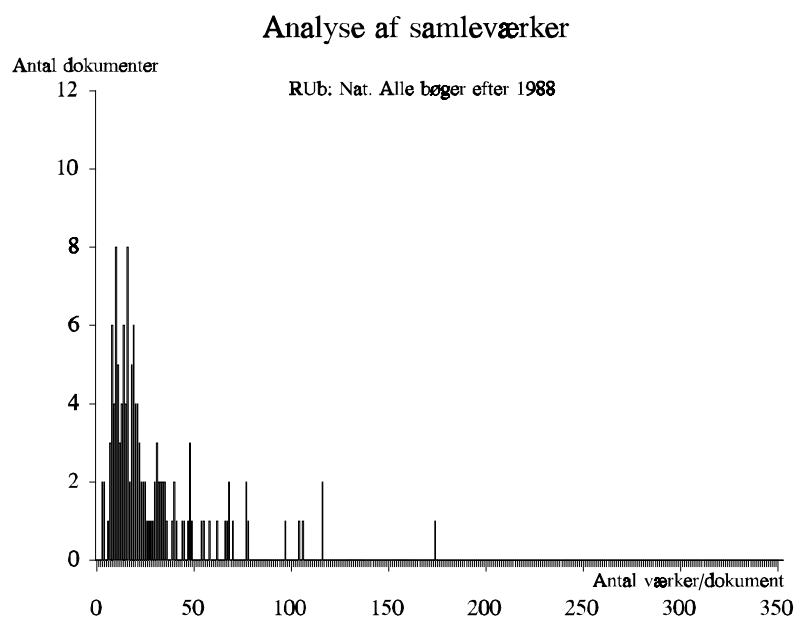
Der blev udtaget dokumenter fra Rub's monografi-katalog indenfor samfundsvidenskab og naturvidenskab, afgrænset ved hjælp af opstillingssignaturen. Dokumenternes publikationsår valgtes at ligge

^a Bredsdorff & Poulsen 1993.

mellem 1988 og 1992. For at få et resultat, der er statistisk sammenligneligt med de hidtidige målinger på DPB's bogmasse udvalgte ca. 700 dokumenter fra hvert fagområde, hvorved der udvalgte hvert tiende dokument inden for de samfundsvidenskabelige grupper, mens det blev hvert tredje dokument inden for de naturvidenskabelige grupper.

Prøven Figur 12 viser at der for de undersøgte 887 bog-poster fra det samfundsvidenskabelige område er 23,7% samleværker. Det betyder at vi med 95% sikkerhed kan sige, at der er $23.7\% \pm 2.9\%$ samleværker i Rub's samfundsvidenskabelige litteratur, publiceret mellem 1988 og 1992.

Prøven viser også, at hvert samleværk i gennemsnit indeholder 16.7 ± 3.1 værker indenfor 95% grænsen.



FIGUR 13

Stikprøveanalyse af andelen af samleværker samt antallet af værker i hvert samleværk. Antal prøver : 698. Antal samleværker: 130. (Bredsdorff & Poulsen 1993)

Prøven Figur 13 viser at der for de undersøgte 698 bog-poster fra det naturvidenskabelige område er 18,3% samleværker. Det betyder

at vi med 95% sikkerhed kan sige, at der er $18,3\% \pm 2,9\%$ samle-
værker i Rub's naturvidenskabelige litteratur, publiceret mellem
1988 og 1992.

Prøven viser også, at hvert samleværk i gennemsnit indeholder $28,0 \pm 4,7$ værker indenfor 95% grænsen.

SAMMENFATNING AF STIKPRØVEANALYSERNE

Resultatet af de seks stikprøveundersøgelser, samlet i Tabel II, viser at antallet af samleværker i gennemsnit udgør ca. 20% på de tre undersøgte biblioteker og at antallet af titler pr. samleværk i gennemsnit på de danske biblioteker er godt 20, mens det er markant højere (ca.30) på det amerikanske college bibliotek.

Bibliotek	Delmængde		Andel Sam- leværker	Antal titler pr. Samle- værk
	Udvalgskriterier	Antal		
Santa Ana		4.098	21,00%	31
DPB	< 1990	718	$16\% \pm 3\%$	20 ± 4
DPB	≥ 1990	211	$17\% \pm 5\%$	17 ± 5
DPB	eng.spr.80/90	496	$15\% \pm 3\%$	29 ± 9
RUb	sam 88/92	887	$24\% \pm 3\%$	17 ± 3
RUb	nat 88/92	698	$18\% \pm 3\%$	28 ± 5

TABEL II

Resultaterne af de 6 refererede stikprøveundersøgelser på Santa Ana College Library, Danmarks Pædagogiske Bibliotek samt Roskilde Universitetsbibliotek.

Går vi lidt mere i detaljer ser vi, at den litteratur, der markant bidrager til at øge antallet af titler i samleværkerne, er den engelsksprogede litteratur. Det kunne forklare det høje antal titler pr. samleværk i undersøgelsen af Rub's naturvidenskabelige litteratur, det markant høje antal titler i undersøgelsen af Danmarks Pædagogiske Biblioteks engelsksprogede litteratur, samt den engelsksprogede litteratur i Santa Ana College Library.

Alt i alt viste de foretagne stikprøver at forskningsbibliotekerne kunne forøge antallet af titler med mindst 400% uden at købe én eneste bog ekstra. Bibliotekerne skjuler altså 4 gange så mange titler som de gør tilgængelige via deres online katalog.

ØKONOMI

Men selvfølgelig koster det resurser at forbedre kataloget. Det koster penge, tid og datalager, at tilføje forfatter og titel for hvert bidrag i et samleværk. Men - og det pointen - det er en marginal omkostning sammenlignet med anskaffelse og sædvanlig systematisering og katalogisering af bogen.

Kravene til datalager er små. Indholdsfortegnelser til samleværker fylder - for et middelstort forskningsbibliotek - omkring 10Mb pr. år^a. Kombineret med de meget lave priser på datalager betyder det, at vi kan se bort fra denne udgift.

Hvad koster det, at tilføje en forfatter og en titel i en allerede eksisterende post?

En berigelse i forbindelse med katalogiseringen kunne ske under de vilkår at bibliotekaren alligevel har posten fremme på skærmen og bogen i hånden. Det er før skønnet, at det tager omkring 20 minutters effektiv arbejdstid at udvælge, korrigere og indskrive en indholdsfortegnelse^b.

Skal denne omkostning sammenlignes med de omkring 3 mandetimer eller 180 minutter, det tog at anskaffe, katalogisere og emnebeskrive en bog (se side 94 ff.) kræver det at de anførte tider

^a I PARADOKS-projektet registreres knap 600 oversigtsværker og marginal litteratur i en database. Af disse er de 294 samleværker, og et udtræk af indholdsfortegnelserne viser, at de, for samleværkernes vedkommende, i gennemsnit fylder 2,06 kb, sammenlignet med Mats Cavallins gennemsnit på 2,12 kb målt på 932 samleværker (Cavallin 1991).

Vi antager på dette grundlag, at indholdsfortegnelsen til et samleværk fylder i gennemsnit to kb, svarende til at hver 1000 nye poster, hvoraf i gennemsnit 200 er samleværker, lægger beslag på 0,4Mb til tekstfilen - med registre omkring 1Mb. For et middelstort forskningsbibliotek, der anskaffer 8.000 nye bøger om året giver det et krav på 8 Mb om året.

^b Atherton 1977, Cochrane 1985.

fordobles, idet den effektive arbejdstid normalt regnes som værende halvdelen af den samlede arbejdstid.

Da omkring 20% af bogmassen har vist sig at være samleværker, vil manuel indtastning af disse samleværkers indholdsfortegnelser i et online katalog udgøre 4% af den samlede arbejdstid^a.

Men kunne det ikke gøres på andre og lettere måder?

Den letteste måde ville være at få indholdsfortegnelsen leveret sammen med litteraturen. Det er allerede tilfældet for dokumenter, der distribueres elektronisk, men en række større boghandlere giver også adgang til at hente indholdsfortegnelserne via en netadgang samtidig med at litteraturen bestilles.

For den store del af litteraturen, hvor en sådan service ikke eksisterer, vil automatiseret indlæsning af indholdsfortegnelser direkte fra bogen være en mulighed.

Fordelen ved skanning og efterfølgende optisk karakterlæsning rapporteres nu som værende meget stor, således at den effektive arbejdstid for indlæsning af en indholdsfortegnelse i gennemsnit bliver lige godt 5 minutter^b.

Alt i alt betyder det, at bibliotekerne kunne financiere en forøgelse af antallet af titler i deres online kataloger med 400% for en omkostning svarende til procestiden for mellem 1 og 4 procent af bøgerne.

^a Når 20% af bogmassen er samleværker, er merarbejdet *i gennemsnit* 20% af 20 minutter - altså 4 minutter. Dette skal sammenlignes med de 180 minutters procestid, svarende til 90 minutter effektivt, hvilket giver et merarbejde på 4%.

^b Selv med anvendelse af en standard PC'er, en standard skanner og et billigt OCR-program reduceres indlæsningstiden væsentligt (I det rapporterede arbejde anvendtes en Macintosh SE/30 med 5Mb RAM samt en Apple skanner med 300 punkts opløsning og Omnipage 2.1 med stavekontrol (Cavallin 1991:9)). Fra de skønnede og hidtil målte 20 minutter ved manuel indtastning reduceredes den effektive arbejdstid til lige godt 5 minutter, fordelt på ca. 60% af tiden til OCR og ca. 40% af tiden til korrektur. Det er vigtigt at bemærke, at der ikke foretages kopiering, men indlæses direkte fra bøgerne, hvilket medfører, at der ikke medgår tid til fotokopiering.

BESKRIVELSE AF PARADOKS-INDEKSERING

PARADOKS-projektet er et forsøg på at skabe et alternativ til det traditionelt emneindekserede bibliotekskatalog.

Udgangspunktet har været det traditionelle katalogs åbenlyse mangler som værktøj til emnesøgning. Disse mangler i det traditionelle katalogs emnebeskrivelser, som PARADOKS-indekseringen skulle udbedre, er behandlet i denne bogs første del, og kan kort opsummeres som

- emneindekseringen er inkonsistent anvendt
- emneindekseringen er foretaget af en for brugeren anonym indekser, hvis perspektiv eller videnskabsforståelse er ukendt for låneren og måske forskellig fra den forfatter, hvis værk indekseres
- emneindekseringen foregår i et for brugeren ukendt kontrolleret sprog
- emneindekseringen er alt for kortfattet til at beskrive en bogs indhold
- emneindekseringen beskriver morgendagens viden med gårsdagens begreber og fungerer således ofte direkte hæmmende for videnskabelige innovationer selv med de bedste intentioner fra indekserens side
- emneindekseringen foretages ikke på artikelniveau for samleværkers vedkommende til trods for, at samleværker repræsen-

terer en udspænding af et emneområde ud fra redaktørens eksplicite viden om emnet, der er kvalificeret sammensat i modsætning til den fragmentarisk sammensatte emnesøgning i et bibliotekskatalog.

Men der er også en anden baggrund for dette projekt.

Hæderkronede professionelle kvaliteter som håndbogskendskab, kendskab til oversigtslitteraturen og den helt usædvanlige litteratur var nødvendige i kortkatalogets dage. Man genbrugte fagligt kvalificerede menneskers emnesøgninger, publiceret i denne litteratur. Man kendte indholdet af litteraturen og havde øje for det, der var anderledes eller marginalt.

I stedet er vi nu ved at få biblioteker, hvor såvel bruger som personale, presset af knappe resurser og manglende tid, benytter bibliotekernes kataloger til emnesøgning - hvilket vi har set ikke altid giver en acceptabel kvalitet.

PARADOKS-projektets intentioner er i høj grad også at højne de professionelle kvalifikationer - ikke på trods af, men i kraft af en fornuftig anvendelse af ny teknologi.

UDFORMNING AF PARADOKSINDEKSERING

For at udbedre de problemer, der ligger i emnesøgning ved hjælp af klassifikationssystemer og kontrollerede emneordssystemer, benytter PARADOKS-indeksering dokumenternes indholdsfortegnelser eller deres abstracts, eventuelt suppleret med andre "naturlige indekser" fra dokumenterne som smudsomslagenes tekst, stikordsregistre, dele af selve teksterne eller centrale referencer^a. Dette er en SAP-lignende indeksering, men uden SAP-indekseringens regler om, at de valgte tekstelementer skal referere til et bestemt antal sider i selve dokumentets tekst.

Som det har været påpeget tidligere, vil en så drastisk forøgelse af antallet af søgbare ord i dokumentbeskrivelserne bevirke en voldsom forøgelse af antallet af fund. Så hvis denne emneindeksering fjerner problemerne ved kontrollerede emnedatasystemer, tilføjes der til gengæld en række nye problemer, først og fremmest

^a Der henvises til Appendiks 3, hvor der er givet nogle eksempler på udformningen af sådanne poster.

at brugerne vil få store problemer med at bruge resultaterne af deres emnesøgninger til noget, da de i forvejen alt for store søgesæt, der er et resultat af traditionel emnesøgning i store bibliotekskataloger, ofte gøres endnu større. Kun for meget specifikke emnesøgninger eller højt specialiserede samlinger uden for mange dokumenter vil et sådant katalog kunne fungere.

For at reducere disse meget store søgesæt og for at kvalificere emnesøgning i bibliotekskataloger, indskrænkes den del af litteraturen, der emneindekseres. Denne reduktion består i at kun oversigtslitteraturen, herunder samleværkerne, samt den marginale litteratur indekseres.

Herved drager man nytte af de emnesøgninger, der allerede er foretaget af oversigtslitteraturens forfattere, og samleværkernes tematisk organiserede artikler bliver søgbare. Den marginale litteratur tilføjes, dels fordi denne litteratur må antages ikke at være særlig rigt repræsenteret i oversigtslitteraturen, og dels fordi denne litteratur repræsenterer et vigtigt potentiale for fremtidige innovationer.

1. PARADOKS-kataloget på Danmarks Pædagogiske Bibliotek

For at kunne foretage en bedømmelse af, hvorvidt et PARADOKS-indeksert katalog er mere eller mindre egnet til emnesøgning end det traditionelle bibliotekskatalog, er det nødvendigt, at foretage en dobbelt registrering af et biblioteks anskaffelser. Derefter kan der foretages sammenligning af emnesøgningsmulighederne i de to katalogudsnit.

På Danmarks Pædagogiske Bibliotek er der foretaget en dobbelt registrering i perioden 1987-1990 af litteratur, publiceret i årene 1985 til 1989 inden for fagene "undervisningsmedier og deres pædagogik" samt "undervisning i de naturvidenskabelige og tekniske fagområder". Registrering af anskaffelser, foretaget i årene 1985-86 blev gjort retrospektivt.

Da brugerne normalt blot er interesseret i dokumenter om deres emne og normalt er ligeglade om der er tale om bøger eller tidsskrifter, indlemmedes såvel bøger som tidsskrifter i det PARADOKS-indekserede katalog. Dette har senere vist sig at møde så store resurseproblemer, at det var strategisk fornuftigt også at undersøge mulighederne for og kvaliteterne af et reduceret PARADOKS-indekseret katalog, udelukkende bestående af bøger.

I perioden 1985-89 indlæstes indenfor de angivne fagområder 1344 poster i det traditionelle katalog, mens der blev registreret 560 PARADOKS-indekseringer. Til trods for det mindre antal poster fylder det PARADOKS-indekserede katalog 30% mere end det traditionelle katalog.^a

Nedenfor beskrives fordelingen af publikationsår og sprog for såvel det PARADOKS-indekserede katalog som for Danmarks Pædagogiske Biblioteks katalog. Endelig beskrives det PARADOKS-indekserede katalogs fordeling af dokumenttyper, samt fordelingen af de kriterier efter hvilke de enkelte dokumenter er udvalgt.

^a I det traditionelle katalogs 1344 poster anvendes i alt 4775 emneordsstreng, svarende til 3,5 emneordsstreng i gennemsnit pr. post. Emneordene fylder i alt 86kb, svarende til, at der i gennemsnit er anvendt ca. 64 karakterer pr. post. Samtidig anvendes 2476 klassemærker, svarende til 1,8 klassemærker i gennemsnit pr. post. Klassemærkerne fylder i alt 17 kb, svarende til ca. 12 karakterer pr. post. I alt fylder det traditionelle katalogs emnebeskrivelse 103 kb, svarende til ca. 76 bytes pr. post.

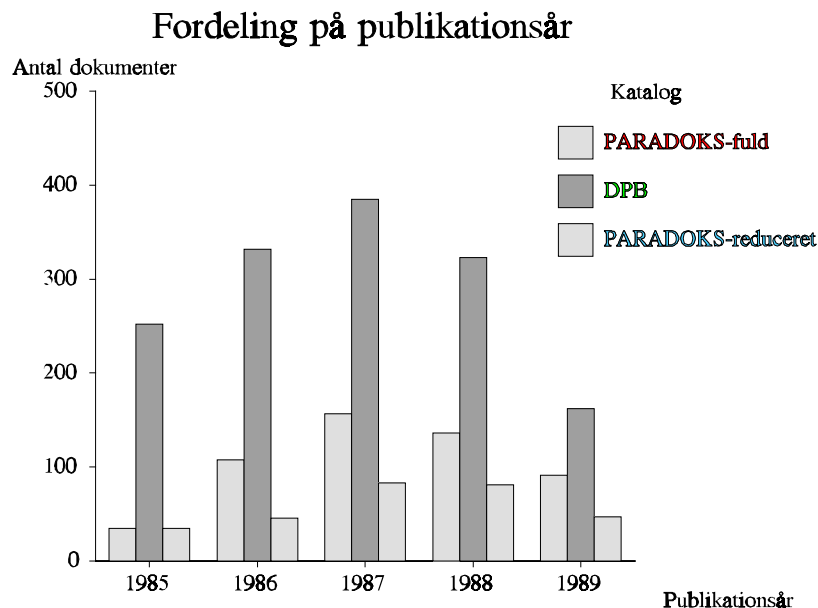
I PARADOKS-katalogets 560 poster udgør emnebeskrivelsen i form af indholdsfortegnelser, resumeer m.v. 764 kb eller 1364 bytes pr. post.

En sammenligning af emneberigelsens størrelse i de to kataloger, viser således, at PARADOKS-katalogets emnebeskrivelse fylder godt syv gange mere end det traditionelle katalogs emnebeskrivelse.

Dette kan umiddelbart virke afskrækkende, men emnebeskrivelsen udgør kun en marginal del af det traditionelle online katalog, hvor hele posten - inklusive emnebeskrivelsen - fylder omkring 0,7 kb. Således fylder det samlede PARADOKS-katalogs tekst $560(0,7+1,4)kb = 1,2Mb$, mens det traditionelle katalogs tekst fylder $1344*0,7kb = 0,9 Mb$. PARADOKS-kataloget fylder altså omkring en tredjedel mere end et traditionelt katalog.

A. PUBLIKATIONSÅR

Den tidlige afgrænsning af dokumenter i begge baser er bestemt af at dokumenterne skal være publiceret i 1985 eller senere og anskaffet på Danmarks Pædagogiske Bibliotek inden 31. december 1989.



FIGUR 14

Fordelingen af publikationsår i det fulde og det reducerede PARADOKS-katalog på baggrund af Danmarks Pædagogiske Biblioteks online katalog.

Figur 14 viser at årene 1985/86 er delvist retrospektivt inddateret til PARADOKS-kataloget, hvilket har betydet lavere udvælgelsesprocenter, men vigtigst var det, at Danmarks Pædagogiske Bibliotek i årene 1987 til 89 intensiverede anskaffelsen af litteratur om anvendelse af interaktive medier i undervisningen samtidig med at denne periode repræsenterede et publikationsboom inden for nævnte emneområde.

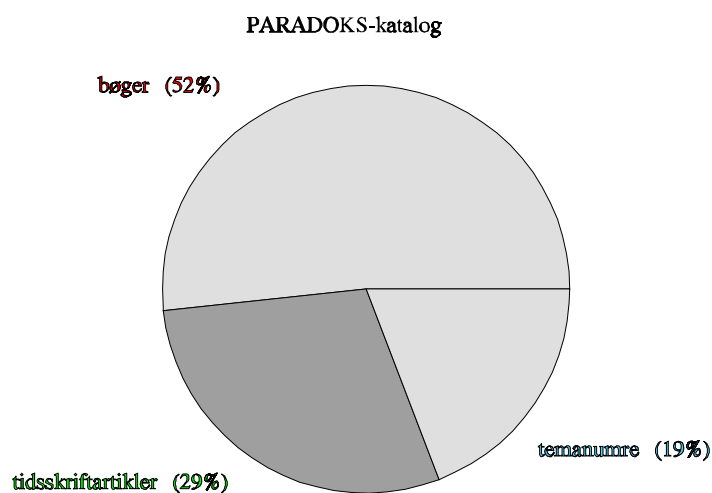
PARADOKS-kataloget reducerer i gennemsnit antallet af poster til 36%, mens det reducerede PARADOKS-indekserede katalog registrerer 20% af posterne.

B. DOKUMENTTYPER

Dokumenter, registreret i Danmarks Pædagogiske Biblioteks online katalog, er monografier. Der bliver i almindelighed ikke foretaget analyser.^a

Dokumenter, registreret i PARADOKS-basen, er dels en delmængde af monografierne i Danmarks Pædagogiske Biblioteks online katalog og dels et udvalg af tidsskriftartikler og alle temanumre indenfor de givne emneområder.

Fordeling af dokumenttyper



FIGUR 15
Fordelingen af PARADOKS-indekseringens dokumenttyper.

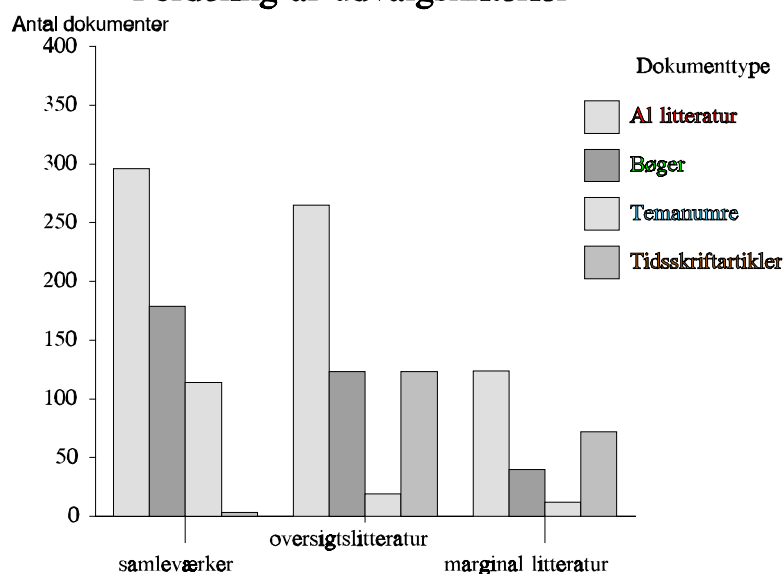
Figur 15 viser, at halvdelen af PARADOKS-kataloget er bøger og den anden halvdel kommer fra tidsskrifter, opdelt således at temanumrene bidrager med ca. 20% og egentlige artikler med ca. 30%.

^a De analyser af danske tidsskriftartikler, der blev foretaget til PEPSY-basen er ikke medtaget.

C. UDVALGSKRITERIER

I modsætning til sædvanlige kataloger, der registrerer alle monografier og tidsskrifternes serietitler, udvælger PARADOKS-registrering kun en del af litteraturen. Der udvælges litteratur ud fra 3 karakteristika: oversigtslitteratur, samleværker og marginal litteratur. Disse kriterier udelukker ikke hinanden, dvs. et dokument kan være udvalgt på grundlag af ét eller flere af kriterierne.

Fordeling af udvalgskriterier



FIGUR 16

Den aktuelle anvendelse af udvælgelseskriterier i det PARADOKS-indekserede katalog, dels for den samlede litteratur og dels fordelt på de forskellige litteratortyper.

Figur 16 viser, at til udvælgelse af den samlede mængde litteratur til PARADOKS-indeksering anvendes kategorierne "samleværker" og "oversigtslitteratur" lige ofte og tegner sig for 82% af de anvendte udvælgelseskriterier, mens "marginal litteratur" tegner sig for 18% af udvælgelseskriterierne.

For bøgernes vedkommende viser det sig, at de for omkring 50% 's vedkommende er udvalgt som "samleværker", omkring 40% er udvalgt som "oversigtslitteratur", og kun 10% af de udvalgte bøger er udvalgt som "marginal litteratur".

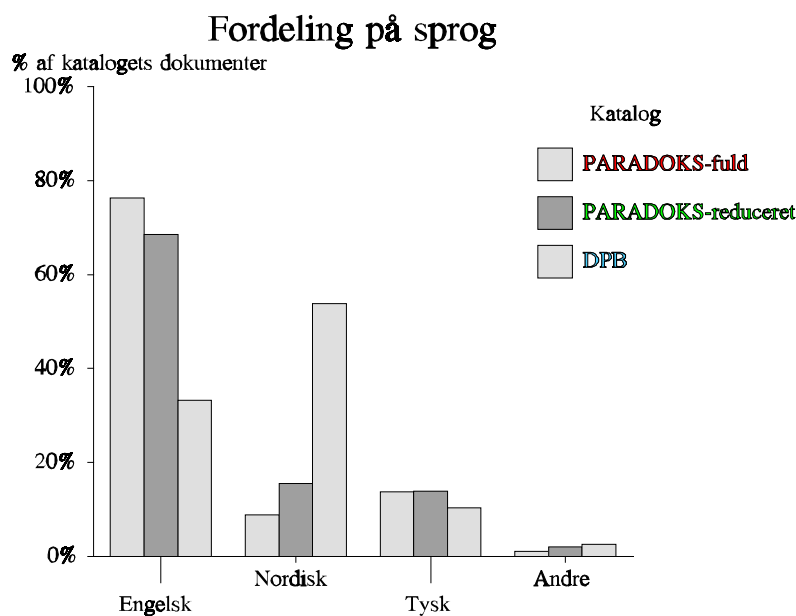
Temanumre er i det væsentlige udvalgt, fordi de netop er samleværker.

Tidskriftartiklerne er for 60%’s vedkommende udvalgt for de er oversigtsartikler, mens hele 40% karakteriseres som "marginal litteratur".

Naturligt nok kommer marginal litteraturen og oversigtslitteraturen fra bøger og tidsskriftartikler, mens samleværker optræder som bøger og temanumre.

D. SPROGLIG FORDELING

Det ses af Figur 17, at PARADOKS-indekseringens udvælgelsesproces favoriserer engelsksproget litteratur, hvad enten det drejer sig om hele PARADOKS-kataloget eller det reducerede PARADOKS-indekserede katalog.



FIGUR 17

Fordeling af dokumenternes nationalsprog i såvel det fulde som det reducerede PARADOKS-katalog med den tilsvarende fordeling i Danmarks Pædagogiske Biblioteks online katalog.

Dette reflekterer, at engelsk er det internationale videnskabelige sprog. Videnskabelige antologier og konferencerapporter skrives

hovedsageligt på dette sprog, og internationale tidsskrifter er ligeledes på engelsk. Selv dansk videnskabelig litteratur publiceres ofte på dette sprog.

Men den sproglige fordeling med den store vægt på de nordiske sprog viser også, at Danmarks Pædagogiske Bibliotek er hovedfagbibliotek for pædagogik, og således skal anskaffe al litteratur om pædagogik, publiceret i Danmark helt uafhængigt af kvaliteten.

Derfor reflekterer den sproglige fordeling blot det sprog på hvilket oversigtslitteratur, samleværker og marginal litteratur udkommer, når denne litteratur har et videnskabeligt niveau.

EVALUERING AF ET PARADOKS-INDEKSERET KATALOG

Når man skal bedømme om én emneindekseringsform er bedre eller dårligere end en anden, må man forsøge at simulere en anvendelsessituation og bedømme resultaterne.

Men en anvendelsessituation er meget kompleks, idet den ikke blot afhænger af emneindekseringen, men også af hvorledes indekseringen præsenteres og hvorledes brugeren anvender kataloget i forhold til egne interesser og egen viden.

Vi har allerede argumenteret for, at der bør satses på at forbedre selve emneindekseringen i online katalogerne. Det er ikke nok, blot at hjælpe brugeren til at udnytte den eksisterende emneindeksering. PARADOKS-indekseringen er netop udviklet med dette for øje, og det er derfor i første omgang relevant at evaluere PARADOKS-indekseringen ud fra emneindekseringens indhold og ikke ud fra emnepræsentation, brugervenlighed m.v..

Litteraturen i et PARADOKS-indekseret katalog består, som det er beskrevet i det foregående kapitel, dels af oversigtslitteratur i bred forstand og dels af marginal litteratur.

De to litteraturtyper har meget forskellig funktion i en emnesøgning med et PARADOKS-indekseret katalog.

Registrering af den marginale litteratur har en traditionel funktion i den forstand, at det PARADOKS-indekserede katalog giver mulighed for at finde den marginale litteratur *direkte*. Søgningen af den marginale litteratur er en ét-trins proces, der fører direkte til litteraturen via en SAP-lignende emneindeksering.

Helt anderledes fungerer PARADOKS-indeksering af oversigtslitteraturen. Her opdeles emnesøgningen i to trin.

Første trin er en søgning efter oversigtslitteratur, der handler om det ønskede emne. Andet trin består i at gennemse den fundne oversigtslitteratur, for at genbruge de emnesøgninger oversigtslitteraturens forfatter allerede har foretaget.

De to forskellige emnesøgningsprocesser resulterer i to forskellige evalueringer, nemlig evaluering af emnesøgning i den marginale litteratur og i oversigtslitteraturen.

I et PARADOKS- indekseret katalog består den marginale litteratur dels af bidrag fra tidsskrifter og dels fra bøger. Begge dele beskrives ved hjælp af indholdsfortegnelser, abstracts samt eventuelt dele af dokumentets tekst.

I et traditionelt bibliotekskatalog registreres kun de marginale *bøger*, og det sker udelukkende ved hjælp af emneord og klassifikationskoder. Det er tidligere beskrevet at kontrollerede indekseringer har store vanskeligheder med at følge med nye udviklinger og usædvanlige emner og angrebsvinkler. Derved er der argumenteret for, at den marginale litteratur er bedre repræsenteret i et PARADOKS- indekseret katalog end i et traditionelt emneindekseret katalog.

Med dette argument vil vi i det følgende udelukkende beskæftige os med den del af det PARADOKS-indekserede katalog, der beskriver oversigtslitteraturen.

Evalueringen kommer således til at foregå som en sammenligning mellem emnesøgningsmulighederne i Danmarks Pædagogiske Biblioteks traditionelt emneindekserede katalog og det parallelt udviklede PARADOKS- indekserede katalog over oversigtslitteraturen i Danmarks Pædagogiske Biblioteks anskaffelser.

Først skal de to kataloger, der skal sammenlignes beskrives nøjere.

DANMARKS PÆDAGOGISKE BIBLIOTEKS KATALOG

Det traditionelle katalog på Danmarks Pædagogiske Bibliotek foretager emneindeksering ved anvendelse af emneord og klassifikation. Som før beskrevet undersøger vi udelukkende de 1344 poster, der repræsenterer dokumenter, publiceret i årene 1985 til

1989 indenfor emneområderne "undervisning i de naturvidenskabelige og de tekniske fag" samt "anvendelse af undervisningsmedier".

I denne periode anvendtes et lokalt udviklet klassifikationssystem, udviklet på basis af DK2 med specielt henblik på pædagogisk litteratur.

Emneordene stammer fra to forskellige kilder - dels fra et lokalt produceret system, beregnet på at beskrive emner, der er tværgående i forhold til det systematiske system - og dels fra den fælles europæiske thesaurus EUDISED, beregnet på en total emnebeskrivelse af dokumenter indenfor pædagogisk forskning.

Mens hvert dokument tildeles mindst én klassifikationskode, er tildelingen af emneord ikke obligatorisk for den betragtede dokumentmasse.

I alt er der i de 1344 poster anvendt 4775 emneordsstreng og 2476 klassifikationskoder, svarende til i gennemsnit 3,5 emneordsstreng og 1,8 klassifikationskode pr. post.

OVERSIGTSLITTERATUREN I DET PARADOKS-INDEKSEREDE KATALOG

Parallelt med registreringen af dokumenter til det traditionelle katalog, blev der foretaget en PARADOKS-indeksering indenfor samme emneområde og samme publikationsår af Danmarks Pædagogiske Biblioteks anskaffelser.

I alt blev der, som beskrevet i forrige kapitel, foretaget 560 PARADOKS-indekseringer, bestående af 289 bøger og 271 bidrag fra tidsskrifter. Oversigtslitteraturens bidrag er 495 poster, bestående af 278 bøger og 217 bidrag fra tidsskrifter.

OVERBLIK OVER EVALUERINGENS STRUKTUR

En sammenligning mellem emnesøgningsmulighederne i ovenstående kataloger er umiddelbart kompliceret, fordi det traditionelle katalog har mange poster, udelukkende bestående af bøger, mens det PARADOKS-indekserede katalog har færre poster, bestående af dels bøger og dels bidrag fra tidsskrifter.

Derved bliver de hyppigste evalueringsmetoder irrelevante, idet sammenligning af rent kvantitative mål som emnesøgningens fuldstændighed og præcision på baggrund af relevansvurderinger, ingen mening har. For hvad er at foretrække: 1 oversigtsværk fra

det PARADOKS-indekserede katalog eller 300 titler fra det traditionelle katalog?

Der er i bogens første del argumenteret for, at et søgeresultat, bestående af færre overbliksgivende titler med større præciseringsmuligheder i form af mere beskrivende tekst, er at foretrække frem for de mange kortfattet beskrevne titler. Men om én titel er nok i det konkrete tilfælde er en rent kvalitativ afgørelse, der afhænger af brugerens vurdering.

Sammenligningen kompliceres yderligere af, at der er en grundlæggende forskel på de to katalogers emnebeskrivelse, idet emnebeskrivelsen i det PARADOKS-indekserede katalog foregår ved hjælp af - i det væsentlige - indholdsfortegnelser og abstracts, mens det traditionelt indekserede katalog benytter sig af klassifikation og emneord. Dette giver et problem ved sammenligning af emnesøgning i de to kataloger, da der naturligvis er forskel på, om der søges på en klassifikationskode h.h.v. et emneord, eller om der søges på nogle tilfældige ord i et dokument.

Men i virkeligheden - d.v.s. i den situation, hvor brugeren foretager en emnesøgning - har vi set, at brugerne anvender ordsøgninger ved at søge i den naturlige tekst^a. Og omvendt er ordsøgning i et PARADOKS-indekseret katalog ikke en søgning på tilfældige ord i et dokument, men en søgning på - i det væsentlige - indholdsfortegnelser og abstracts i den overbliksgivende litteratur og samleværker. For oversigtslitteraturens vedkommende er der ikke tale om tilfældige tekstudvalg, men derimod forfatterens egen indeksering af sin tekst. For samleværkernes vedkommende er indholdsfortegnelsens ord en indgang til samlingens emne, og således betydelig mere værdifuld end et titelord i det traditionelle online katalogs samling af titler, der kun har det til fælles, at de viser, hvad biblioteket har anskaffet.

Derfor er forskellen i praksis ikke stor.

Med denne baggrund kan vi opstille en struktur for sammenligningen mellem de to katalogers emnesøgningmuligheder.

^a Hvis katalogets søgesystem er indrettet således, at basisindekset også indeholder emneordsfelterne, kan brugeren *tilfældigvis* ramme et emneord. Dette er naturligvis ikke en emnesøgning i almindelig forstand.

FINDES EMNEORDENE I DET PARADOKS-INDEKSEREDE KATALOG?

Først undersøges det, om de anvendte emneord i det traditionelle katalog også findes i det PARADOKS-indekserede katalog over oversigtslitteraturen.

Selv om al den undersøgte litteratur er klassificeret og opstillet som litteratur om "undervisning i de naturvidenskabelige og de tekniske fag" samt "anvendelse af undervisningsmidler", handler det beskrevne dokument ofte om aspekter af disse fag, som det er emneordssystemets opgave at beskrive. Det er derfor relevant at se, om man ved anvendelse af disse emneord også finder oversigtslitteratur igennem det PARADOKS-indekserede katalog.

FINDES KLASSEFIKATIONSKODERNE I DET PARADOKS-INDEKSEREDE KATALOG?

Dernæst undersøges det, om de anvendte klassifikationskoder i det traditionelle katalog også optræder i det PARADOKS-indekserede katalog i deres sproglige ækvivalent.

I modsætning til, hvad der er tilfældet for anvendelse af emneord, tildeles al litteratur i det traditionelle katalog en klassifikationskode. Derved bestemmes en emneafgrænsning, der anvendes ved dokumenternes opstilling og ved denne undersøgelses emneafgrænsning. Undertiden tildeles et dokument flere klassifikationskoder - også koder, der ligger udenfor denne undersøgelses emneafgrænsning. I evalueringen medtages kun de klassifikationskoder, der ligger indenfor undersøgelsens emnemæssige afgrænsning, idet det PARADOKS-indekserede katalog er konstrueret primært til at dække dette emneområde.

HVOR ER PRÆCISERINGSMULIGHEDERNE BEDST?

Nu anvendes klassifikationkoder ikke som det eneste værktøj til at finde litteratur om et emne.

Først og fremmest fordi antallet af fund ved anvendelse af klassifikationskoder - som vi har set tidligere - ofte er så stort, at resultatet bliver uanvendeligt for brugeren.

Dernæst fordi det normalt giver dårlige resultater, at foretage yderligere begrænsninger af et søgesæt ved hjælp af klassifikationskoder eller emneord. Årsagen er, at en logisk "både og" kombination af to meget inkonsekvent og inkonsistent anvendte indek-

seringer med stor sandsynlighed giver endnu dårligere resultat, end hvis disse indekseringer kun var anvendt én gang.

Derfor kombineres en klassifikationsbaseret emnesøgning næsten altid med en søgning på naturlig tekst i større online kataloger - om ikke andet ved den efterfølgende gennembladning af søgesættets fund.

Heraf følger, at en sammenligning af emnesøgningsmulighederne i de to kataloger, kræver en uddybende undersøgelse af hvilke specificerende søgninger med naturlig tekst, der er mulige indenfor de søgesæt, den grove emnesøgning med klassifikationskoder frembragte.

REPRODUCÉRBAR EVALUERINGSMETODE

Metoden udføres således, at udvælgelsen af emneord og klassifikationskoder samt selve analysen er reproducérbar. Hermed menes, at det tilstræbes, at evalueringen er uafhængig af den, der foretager evalueringen, og der redegøres eksplicit for hvert af undersøgelsens trin.

Vi undgår hermed - i princippet - det stærkt subjektive element, der ligger i udvælgelsen af hvilke cases, der undersøges.

Vi undgår også, at den lidt kunstige problematik, der består i spørgsmålet, om hvad der er bedst: Ét oversigtsværk eller tre hundrede "almindelige bøger"?, bliver alt for central, idet præcise fund er at foretrække frem for megen støj.

1. Overflødiggor PARADOKS-indeksering emneord i online kataloger?

Her vil vi undersøge, om de emneord, der eksisterer i det traditionelle katalog, også optræder i det PARADOKS-indekserede katalog^a over oversigtslitteraturen.

For at komme anvendelsessituationen så nær som muligt, oversættes det traditionelle katalogs emneord "råt" til det relevante sprog, og det undersøges, om disse emnebeskrivelser kan genfindes i den PARADOKS-indekserede base.

Det betyder - groft taget - at hvis PARADOKS-kataloget indeholder emneordene eller deres oversættelser, tilbyder PARADOKS-kataloget de samme søgemuligheder, som det traditionelle online katalog. Omvendt betyder en manglende genfindning af emneordene i PARADOKS-posterne, at denne emnesøgningsmulighed ikke findes i det PARADOKS-indekserede katalog.

For at undersøge nødvendigheden af tidsskrifters bidrag til emnesøgningsmulighederne i et PARADOKS-indekseret katalog undersøges dels det fulde PARADOKS-indekserede katalog, og dels det reducerede PARADOKS-indekserede katalog, der er begrænset til bøger alene.

Af praktiske årsager udvælges kun de mest anvendte emneord til undersøgelse.

Det kan delvist retfærdiggøres af, at en af de centrale styrker ved at anvende kontrollerede sprog er, at hvert emneord samler litteratur, der ellers er beskrevet med forskellige synonymer og på forskellige sprog. Hvis emneordet da kun anvendes i yderst få tilfælde i den undersøgte periode bortfalder den nævnte fordel for de kontrollerede sprog.

^a I litteraturen er der foretaget adskillige undersøgelser af om emneordene allerede skulle forekomme i den *tilsvarende* titel (Henzler 1978, Maizell 1960, Ruhl 1964, Montgomery & Swanson 1962, Bottle 1970). I evalueringen af det PARADOKS-indekserede katalog er det vigtigt at indse, at disse resultater ikke kan benyttes, idet det traditionelle katalog og PARADOKS-kataloget ikke har samme dokumentmasse.

DANMARKS PÆDAGOGISKE BIBLIOTEKS ONLINE KATALOG

Danmarks Pædagogiske Biblioteks online katalog indeholder 1344 poster, produceret i perioden 1985-89 indenfor emneområderne naturvidenskabernes og teknologiernes pædagogik samt undervisningsmedier. Disse 1344 poster blev analyseret for emneord, der optræder 4 gange eller mere.

Analysen gav 157 emneordsstrengene, der havde været brugt 4 gange eller mere i Danmarks Pædagogiske Biblioteks 1344 undersøgte poster.

For at kunne anvende disse emneord til søgning i PARADOKS-katalogets naturlige tekst blev de oversat til engelsk og tysk. Oversættelsen var ganske mekanisk.

Der tilføjedes kun sjældent synonymmer således at det svarer meget godt til en gennemsnitlig brugers grove sprogbehandling i disse situationer. Emneordsstrengene er opført i Appendiks 4.

RESULTATER

Resultaterne af sammenligningen mellem mulighederne for søgning på de allerede anvendte emneord på Danmarks Pædagogiske Biblioteks katalog og de PARADOKS-indekserede kataloger er opstillet i Appendiks 4.

Her finder vi, at ud af de 157 hyppigst forekommende emneord i det undersøgte udsnit af Danmarks Pædagogiske Biblioteks katalog, forekommer de 128 - eller 81% - i det fulde PARADOKS-indekserede katalog over oversigtslitteraturen, mens 127 - eller 80% - fandtes i det PARADOKS-indekserede katalog over oversigtslitteraturens bøger. Det betyder, at hvad enten de 1344 poster i det traditionelt indekserede katalog erstattes med de 495 poster i det fulde PARADOKS-indekserede katalog over oversigtslitteraturen, eller de 278 poster i det PARADOKS-indekserede katalog over oversigtslitteraturens bøger, så er 4 ud af 5 emneord i det traditionelle katalog til stede i det PARADOKS-indekserede katalog over oversigtslitteratur.

DISKUSSION

Den høje forekomst af emneord i de PARADOKS-indekserede kataloger skal ses på baggrund af, at mange af emneordene er af en

sådan art, at de ikke kan forventes at optræde i naturlig tekst. Eksempelvis ligger det implicit i en englænders beskrivelse af skolesystemet, at det drejer sig om det engelske skolesystem, hvis ikke andet er anført. Andre emneord er af administrativ art - "fag i læseplan" anvendes om alle skemalagte fag (matematik i den danske folkeskole, eller teknik efter 1994, men ikke teknik før 1994, hvor det blev indført som fag).

Andre emneord er ikke oversættelige til naturlig tekst - for eksempel "informationsteknologi, informationssamfund og uddannelse".

Visse emneord giver meget lidt relevant materiale - således "forsøg", der benyttes til at betegne forsøg med undervisningen og ikke den undervisningsmetode, der anvender forsøg, f.eks. fysikforsøg.

Derfor må det konkluderes, at kvaliteten af emnesøgning med emneord i det PARADOKS-indekserede katalog er højest for emneord, der *ikke* betegner geografiske, administrative og visse dokumenttypologiske kategorier.

2. Overflødiggør PARADOKS-indeksering systematisk klassifikation i online kataloger?

Her vil vi undersøge, om den systematiske klassifikation, der eksisterer i det traditionelle katalog, også optræder i det PARADOKS-indekserede katalog over oversigtslitteraturen med sin sproglige ækvivalent eller klassens navn.

Som før oversættes de anvendte klassers navne "råt" til de relevante sprog og kun de mest anvendte systematiske klassifikationskoder undersøges.

DANMARKS PÆDAGOGISKE BIBLIOTEKS ONLINE KATALOG

Danmarks Pædagogiske Biblioteks online katalog indeholder 1344 poster, produceret i perioden 1985-89 indenfor emneområderne naturvidenskabernes og teknologiernes pædagogik samt undervisningsmedier. Disse 1344 poster blev analyseret for systematiske klassifikationkoder, der optræder 4 gange eller mere.

Analysen gav 81 klassifikationskoder, der havde været brugt 4 gange eller mere i Danmarks Pædagogiske Biblioteks 1344 undersøgte poster. Af disse 81 klassifikationskoder hører 37 til det undersøgte emneområde. Disse 37 beskrivelser af de systematiske klassifikationskoder er oversat til engelsk og tysk og er opført i appendiks 5.

RESULTATER

Resultaterne af sammenligningen mellem mulighederne for søgning på de allerede anvendte klassifikationskoder på Danmarks Pædagogiske Biblioteks katalog og de PARADOKS-indekserede kataloger er opstillet i appendiks 5.

Her finder vi, at ud af de 37 hyppigst forekommende klassifikationskoder i det undersøgte udsnit af Danmarks Pædagogiske Biblioteks katalog, forekommer de 29 - eller 78% - i det fulde PARADOKS-indekserede katalog over oversigtslitteraturen, mens 26 - eller 70% - fandtes i det PARADOKS-indekserede katalog over oversigtslitteraturens bøger.

Det betyder, at hvad enten de 1344 poster i det traditionelt indekserede katalog erstattes med de 495 poster i det fulde PARADOKS-indekserede katalog over oversigtslitteraturen eller de 278 poster i det PARADOKS-indekserede katalog over oversigtslitteraturens bøger, så er 3 ud af 4 af det traditionelle katalogs systematiske klassifikationskoder til stede i det PARADOKS-indekserede katalog over oversigtslitteratur.

DISKUSSION

Resultatet for de systematiske klassifikationskoder er lidt dårligere, end resultaterne for emneordssøgningerne i forrige kapitel, men der er stadig en høj grad af sammenfald.

Årsagerne til det dårligere resultat skal først og fremmest søges i, at overordnede systematiske klassifikationsgrupper ikke repræsenterer dokumenterne eller litteraturen, men bibliotekernes eller systemets dannelse af overordnede begreber. "Anvendt videnskab" vil altid være mere specifikt repræsenteret i dokumenterne - og vel også hos forskere og studerende, der sjældent studerer anvendt videnskab, men nærmere bærende konstruktioner eller vindmøller. Disse overordnede og systembestemte systematiske grupper er

anvendelige til emneafgrænsning af litteraturen før egentlig emnesøgning foretages og bør - bl.a. i konsekvens af disse gruppers ringe repræsentation i de PARADOKS-indekserede kataloger - kombineres med en sådan indeksering.

Andre svagt repræsenterede grupper er fag som "Husvæsenets økonomi", "Levnedsmiddelindustri" eller "Vedvarende energi". Alle disse grupper er enten meget "små" fag eller meget "lokale" fag i forhold til det aktuelle biblioteks anskaffelsespolitik. Det bør som konsekvens heraf overvejes hvorledes disse fag tilgodeses i forbindelse med PARADOKS-indekserede kataloger.

3. Sammenligning af præciseringsmuligheder i et PARADOKS-indeksret og et traditionelt online katalog

Vi har lige erfaret, at man ved søge på emneord eller klassifikationernes standardbeskrivelser i PARADOKS-kataloger kun mister én ud af fire til fem svarmuligheder.

Med den svage anvendelse af disse emnedatasystemer og disse systemers ringe muligheder for emnebeskrivelse er dette tab ikke et problem. Problemet ligger i, om det er muligt at lave en mere præcis og først og fremmest forståelig form for emnebeskrivelse, der samtidig giver en god kvalitet i emnesøgningen.

Med den stigende mængder af publikationer er det mulighederne for at foretage en relevant præcisering af emnesøgningen, der er det centrale krav.

Vi skal her sammenligne præciseringsmulighederne i det traditionelle online katalog på Danmarks Pædagogiske Bibliotek og et tilsvarende PARADOKS-indekserede katalog.

UDVÆLGELSE AF EMNESØGNINGER TIL UNDERSØGELSE

Det vil være helt uden interesse at foretage en bred sammenligning af præciseringsmulighederne mellem det traditionelle emneindekserede katalog, og et tilsvarende PARADOKS-indekserede katalog.

Antallet af ord i et PARADOKS-indekserede katalog er simpelthen større end antallet af ord i det traditionelt indekserede katalog^a.

I det følgende vil vi derfor sammenligne præciseringsmulighederne indenfor søgesæt, dannet ved klassifikationssøgning. For at søgesæt, dannet i et PARADOKS-indekserede katalog over oversigtslitteraturen er sammenligneligt med det traditionelle kræves det, at de fundne dokumenter i det traditionelle katalog er hovedplaceret i en systematisk gruppe, der ligger indenfor det afsøgte emneområde.

Vi vil nu søge at opstille en metode til at sammenligne præciseringsmulighederne indenfor disse søgesæt.

HVORLEDES MÅLES PRÆCISERINGSMULIGHEDER?

Som nævnt forstår vi her præciseringsmuligheder som de muligheder, søgesættens naturlige tekst giver for at specificere et emne indenfor den afgrænsning, der er foretaget med en klassifikations-søgning.

Opgaven består i at bestemme hvor mange "relevante" ord, der forekommer i de to søgesæt, stammende fra klassifikationssøgning i det traditionelle katalog og søgning med klassifikationens betegnelse i de PARADOKS-indekserede kataloger over oversigtslitteraturen.

RELEVANTE ORD

Nu er ikke alle ord lige relevante, hverken i det PARADOKS-indekserede katalog eller i det traditionelle online katalogs titler. For at bedre på denne situation burde kun de ord medtages, der findes i den aktuelle sprogbrug i det aktuelle problemfelt. I denne undersøgelse er ord, der forekommer i nye artiklers abstracts, valgt som de relevante ord inden for emnet.

Til trods for, at kun ord i abstracts fra nyere artikler om emnet tælles med, er der stadig ord, der ikke kan anvendes som præciseringsmulighed. En stor del af de mindre anvendelige ord frasorteres ved anvendelse af en standardiseret liste af stopord. Men stadig må disse målinger ikke tolkes for rigoristisk og fordobling af

^a Se fodnote a side 108.

antallet af ord betyder nok en betydelig forøgelse af præciseringsmulighederne, men ikke nødvendigvis en fordobling af disse. Med dette udgangspunkt undersøges hvilken forøgelse eller reduktion af søgemulighederne, PARADOKS-kataloget repræsenterer i forhold til det traditionelle katalog.

FORSKELLIGE ENDELSER AF SAMME ORD

Hvis nu det ene katalog indeholder samtlige bøjningsformer for de indgående ord, mens det traditionelle katalog indeholdt ganske få forskellige endelser på de indgående ord, ville sammenligningen være mindre værdifuld til belysning af mulighederne for at præcisere emnesøgninger. Hvad enten der står "geological" eller "geology" i en titel eller en indholdsfortegnelse, er det rimeligt at antage, at dokumentet handler om geologi og de to ord repræsenterer altså kun én præciseringsmulighed.

Derfor undersøges det, om resultatet skulle være følsomt over for forskellige endelser. Dette gøres ved at gentage sammenligningerne, men kun for de første 9 h.h.v. 7 og 5 bogstavers vedkommende.

HVILKE DELE AF POSTERNE SAMMENLIGNES?

Men hvad er det egentlig, der sammenlignes?

Det traditionelle katalogs bidrag til den naturlige tekst består i titler, undertitler, forfatter(e), serietitel, forlag, sted, publikationsår m.v.. I det følgende kaldes det den bibliografiske identifikation (BI).

PARADOKS-katalogets bidrag til den naturlige tekst består normalt i et mindre antal poster end det traditionelle katalog, men til gengæld indeholder hver post, ud over den bibliografiske identifikation, også indholdsfortegnelse, evt. abstract eller andre indholdsbeskrivende dele af dokumentet.

Mens det traditionelle katalogs bibliografiske identifikation antages at være relevant i forhold til emnesøgningen, behøver det ikke at gælde den PARADOKS-indekserede posts indholdsfortegnelse eller abstract.

Eksempel

Hvis et samleværk om danskerne i slutningen af det tyvende århundrede indeholder bidrag om kvindefrigørelse og et bidrag om naturvidenskabernes udvikling, er det ikke sikkert, at dokumentet er relevant for emnet "kvinder og naturvidenskab".

Hvis det derimod kræves, at såvel kvinder som naturvidenskab står i samme bidrag, vil det være relevant for emnet.

Havde værkets titelord imidlertid indeholdt såvel kvinder som naturvidenskab, er det rimeligt at antage, at alle bidragene er relevante for emnet.

Derfor inddrages kun dele af den PARADOKS-indekserede post i sammenligningen, nemlig de dele, der må antages at være relevante for den forudgående grove emnesøgning.

Derfor udførtes alle søgninger i PARADOKS-kataloget ved hjælp af nærhedsoperatorer, der sikrer at den samlede emnestreng forekommer i samme felt eller endnu tættere. Derved reduceres muligheden for at ord, der er uden meningsmæssig sammenhæng, kommer til at resultere i et fund ved søgning med pågældende emnestreng.

Og derfor medtages hele den PARADOKS-indekserede post kun i de tilfælde, hvor hele emnestrengen findes i den bibliografiske identifikation.

SPROG

Undersøgelsen begrænses til kun at medtage engelsksprogede dokumenter i såvel det traditionelle som i det PARADOKS-indekserede katalog over oversigtslitteraturen.

Valget er praktisk begrundet. Samtidig gælder det for begge katalogers vedkommende, at dokumenter, der er interessante for emnesøgning, er affattet på engelsk^a.

Denne sproglige afgrænsning betyder en reduktion i antallet af dokumenter, der skal sammenlignes.

Vi fandt før, at ud af de 37 klassifikationskoder, der var anvendt 4 gange eller mere i de 1433 poster fra det traditionelle katalog, var 29 repræsenteret via klassifikationskodens navn i det fulde PARADOKS-indekserede katalog. Det viser sig, at 4 ud af de 29 klassifikationskoder, der også fandtes i det PARADOKS-indekserede katalog, ikke anvendtes til beskrivelse af den engelsksprogede litteratur på Danmarks Pædagogiske Bibliotek i det undersøgte tidsrum. Tilbage er 25 klassifikationskoder, der er anvendt på engelsksproget

^a Se beskrivelsen af den sproglige fordeling af dokumenter i PARADOKS-kataloget side 112.

litteratur og som findes i det tilsvarende PARADOKS-indekserede katalog over oversigtslitteraturen. Disse koder er opført i appendiks 6.

TEKNISK BESKRIVELSE

For hver af de 25 søgeprofiler dannes resulterende søgesæt fra Danmarks Pædagogiske Biblioteks traditionelt indekserede katalog, fra det fulde PARADOKS-indekserede katalog og fra det reducerede PARADOKS-indekserede katalog, udelukkende bestående af bøger. Og endelig dannes et "relevans"-søgesæt ved en søgning i online tidsskriftkataloget "ERIC/CIJE", hvor kun sættets abstracts medtages.

Søgesættet fra Danmarks Pædagogiske Bibliotek, fra PARADOKS-indekserede kataloger og "ERIC/CIJE" lægges ind i databaseprogrammet "brs-search", hvorefter ordlisterne fra de tre tekstfiler trækkes ud. Fællesmængden mellem ERIC/CIJE-listen og h.h.v. Danmarks Pædagogiske Bibliotek og det PARADOKS-indekserede katalog dannes. Derved dannes to lister af "relevante" PARADOKS-ord og DPB-ord.

Til sidst undersøges det i hvilken udstrækning det PARADOKS-indekserede katalog indeholder det traditionelt indekserede DPB-katalogs ord ved at danne de to residualmængder og fællesmængden mellem de relevante DPB-ord og de relevante PARADOKS-ord.

Processen gentages for ordlængderne 9, 7 og 5 bogstaver.

Alle processerne er programmeret sammenhængende i Word Perfect's og Word Perfect Shell's makrosprog, hvori de nødvendige optællinger også er foretaget.

RESULTATER

Resultaterne af de 25 sammenligninger mellem præciseringsmulighederne i det traditionelt indekserede katalog på Danmarks Pædagogiske Bibliotek og det PARADOKS-indekserede katalog er angivet i appendiks 6 hvoraf det fremgår, at alle resultaterne er stabile overfor de indgående ords endelsesvariationer. Resultaterne er sammenfattet i Tabel III. Alle sammenligninger er dubleret således, at der både sammenlignes med det fulde PARADOKS-indekserede katalog og den del, der kun indekserer bøger. Der henvises til appendiks 6 for detaljer.

Tabellen angiver antallet af fund i Danmarks Pædagogiske Biblioteks katalog (DPB) og de to PARADOKS-indekserede kataloger. Dernæst anføres antallet af ord fra søgningen i DPB's base, som samtidig optræder i nyere abstracts om emnet, taget fra ERIC.

Nettoresultatet er beregnet i forhold til antallet af relevante ord i DPB-søgningerne. Således betyder 30%, at søgningen i det PARADOKS-indekserede katalog gav poster, der indeholder 30% flere relevante ord end søgningen i Danmarks Pædagogiske Biblioteks katalog. Og omvendt betyder -25%, at søgningen i det PARADOKS-indekserede katalog gav poster, der indeholder 25% færre relevante ord end søgningen i Danmarks Pædagogiske Biblioteks katalog.

Emne	Antal fund			Relevante ord	Nettoresultat	
	DPB	PARADOKS		DPB	PARADOKS	
		Alle	Bøger		Alle	Bøger
Biologi. Metodik	7	14	11	35	274%	151%
Datalogi. Metodik	39	10	6	86	20%	-1%
Matematikvanskeligheder	14	8	8	40	123%	123%
Elektronik og elektroteknik. Metodik	1	6	4	1	3100%	2800%
Energi. Metodik	5	3	2	16	269%	181%
Fysik. Metodik	14	37	26	40	350%	300%
Geografi. Metodik	6	20	9	23	700%	396%
Geologi. Metodik	1	8	2	5	1220%	860%
Husholdning. Metodik	4	1	1	8	-27%	-27%
Datamaskinstyret undervisning	149	71	52	342	87%	76%
Kemi. Metodik	6	15	9	35	286%	246%
Landbrugsundervisning	3	3	1	10	380%	-20%

Emne	Antal fund			Relevante ord	Nettoresultat	
	DPB	PARADOKS		DPB	PARADOKS	
		Alle	Bøger		Alle	Bøger
Madlavning. Metodik	2	4	3	6	250%	233%
Matematik. Metodik	113	143	87	264	264%	238%
Matematik og samfund. Metodik	3	19	11	4	2525%	2225%
Naturvidenskab og samfund	6	3	1	18	39%	-56%
Ernæringslære. Metodik	14	2	2	61	0%	0%
Teknologi og samfund. Metodik	3	26	18	10	860%	750%
Færdselslære. Metodik.	1	3	3	9	156%	156%
Økologi. Metodik	11	26	11	60	533%	402%
Naturkundskab. Metodik	89	14	8	281	4%	-14%
Audiovisuelle hjælpemidler. Metodik.	9	5	1	22	277%	-59%
Mediepædagogik	19	25	10	35	377%	180%
Teksttilvirkning, håndarbejde	1	3	3	4	250%	250%
Byggeteknik, arkitektur	1	3	2	4	75%	50%

TABEL IV

Analyse af præciseringsmuligheder i 20 klassifikationssøgninger, i Danmarks Pædagogiske Biblioteks katalog og i det PARADOKS-indekserede katalog. Nettoresultatet er beregnet i forhold til antallet af relevante ord i DPB-søgningerne. Således betyder 30%, at søgningen i det PARADOKS-indekserede katalog gav poster, der indeholder 30% flere relevante ord end søgningen i Danmarks Pædagogiske Biblioteks katalog. Og omvendt betyder -25%, at søgningen i det PARADOKS-

indekserede katalog gav poster, der indeholder 25% færre relevante ord end søgningen i Danmarks Pædagogiske Biblioteks katalog. Der henvises til Appendiks 6 for nærmere information.

Af tabellen fremgår det, at i 18 af de 25 tilfælde indeholder søgningen i det PARADOKS-indekserede katalog præciseringsmuligheder, der langt overgår det traditionelle katalogs præciseringsmuligheder. Dette gælder for såvel det fulde PARADOKS-indekserede katalog som for det reducerede katalog, hvor kun bøgerne er medtaget. Det gælder endog, hvor antallet af PARADOKS-indekserede poster er betydelig mindre end antallet af poster, fundet i det traditionelle katalog.

De syv resterende søgninger er fordelt således, at to var neutrale ("Datalogi" og "Naturkundskab"), i fire af søgningerne ("Ernæringslære", "Landbrugsundervisning", "Audiovisuelle hjælpemidler" og "Undervisningsfaget naturvidenskab og samfund") havde det fulde PARADOKS-indekserede katalog bedre præciseringsmuligheder end det traditionelle katalog, mens det reducerede katalogs søgninger gav samme eller dårligere præciseringsmuligheder end det traditionelle katalog. Endelig indeholdt søgningen på "Husholdningslære" dårligere præciseringsmuligheder i både det fulde og det reducerede PARADOKS-indekserede katalog end i det traditionelle katalog.

4. Konklusion på evalueringen af PARADOKS-indeksering

Evalueringen blev foretaget som en sammenligning af emnesøgningsfaciliteterne i Danmarks Pædagogiske Biblioteks 1344 poster og de to tilsvarende PARADOKS-indekserede kataloger over den tilsvarende oversigtslitteratur på henholdsvis 495 og 278 poster.

Det har vist sig, at de traditionelle emneindekseringer i Danmarks Pædagogiske Biblioteks katalog i omkring 75% af tilfældene blev genfundet i begge de PARADOKS-indekserede katalogers naturlige tekst, taget fra indholdsfortegnelser eller abstracts fra oversigtslitteratur. Dette sætter endnu et alvorligt spørgsmålstegn ved bibliotekernes kostbare emneindekseringspraksis.

Men indekseringen kunne være nødvendig, hvis præciseringsmulighederne var markant større i det traditionelt indekserede katalog - de præciseringsmuligheder, der gør, at katalogets brugere enten kan foretage en mere præcis søgning i dagens store informationsmængder eller kan foretage en fagligt forsvarlig udvælgelse af relevante dokumenter.

Sammenligning af søgeresultaterne fra klassifikationssøgning i det traditionelt indekserede katalog og de tilsvarende søgninger i de to PARADOKS-indekserede kataloger viste, at i mellem 80% og 96% af søgningerne, havde de PARADOKS-indekserede kataloger forbedrede eller lige så gode præciseringsmuligheder, som det traditionelt indekserede katalog havde. Det bedste resultat opnåedes naturligvis for det fulde PARADOKS-indekserede katalog med 495 poster med både bøger og tidsskrifter repræsenteret. Her var der tale om forbedrede eller samme præciseringsmuligheder i 24 ud af 25 søgninger. For det reducerede PARADOKS-indekserede katalog over oversigtslitteratur, dækkende 278 bøger, havde 20 ud af 25 søgninger forbedrede eller samme præciseringsmuligheder.

Det betyder, at forskellen mellem at indeksere 495 bøger, tidsskriftartikler og temanumre og 278 bøger er overraskende lille.

Herudover peger den meget konservative evalueringsform, hvor der søges med emneord og systematiske klassifikationskoder, at det PARADOKS-indekserede katalogs tilbud om genbrug af de emnesøgninger, der er oversigtslitteraturens grundlag, undervurderes. Enhver søgning med synonymer, meget specifikke emner eller metoder vil give mulighed for at finde endnu mere oversigtslitteratur.

Undersøgelsen peger således på, at man til emnesøgningsformål med fordel kan erstatte de 1344 poster i det traditionelle katalog med et PARADOKS-indekseret katalog på 278 poster, dækkende den del af de 1344 poster over bøger, der referer til oversigtslitteratur.

DISKUSSION

Den præsenterede evaluering er udviklet ud fra et ønske om at distancere sig fra megen evaluering i informationsvidenskaben, der bygger på meget subjektive vurderinger af relevans og først og fremmest i udvælgelse af de spørgsmål, der stilles. I og med, at det ofte er konstruktøren af forbedringerne, der også foretager evalu-

eringen, kan det - selv med de bedste intentioner - give evalueringen en vis slagside.

Her er der tilstræbt en metode, der først og fremmest har sin styrke ved at være reproducerbar. Dette gælder først og fremmest i udvælgelsen af de søgninger, der sammenlignes i de to kataloger. Alle udvælgelser af emner er foretaget på grundlag af emnernes hyppighed i det traditionelle katalog.

Realiseringen af den beskrevne evalueringsmetode har sine begrænsninger.

Oversigtslitteratur er hovedsagelig engelsksproget og evalueringen af præciseringsmulighederne er udelukkende begrænset til engelsksproget litteratur.

I et lille sprogområde, som det danske, er det uheldigt, hvis vi undertrykker emnesøgning i dansksproget litteratur via vores kataloger.

Dette er ikke undersøgt.

På den anden side er det et faktum, at den stigende internationalisering af den videnskabelige debat har gjort engelsk til det helt dominerende i litteraturen.

En sidste begrænsning i evalueringens gyldighed er knyttet til, at der kun er foretaget kvantitativ evaluering af de to kataloger.

Naturligvis er det normalt en mangel ikke at medtage de forskellige emnesøgningers kvalitet. Men denne sammenligning er ikke normal. Der sammenlignes nemlig to kvalitativt forskellige kataloger. Det traditionelle katalog, der medtager alt fra højt specialiserede rapporteringer til oversigtslitteratur. Det PARADOKS-indekserede katalog, der kun medtager oversigtslitteraturen. Derfor vil et fund i det PARADOKS-indekserede katalog give en oversigt over emnets litteratur, mens et fund i det traditionelle katalog normalt blot vil repræsentere et enkelt dokument.

Således dækker den kvantitative evaluering over en kvalitetsforskel med henblik på emnesøgning, der betyder, at samme antal emnesøgningssmuligheder i de to kataloger normalt giver den bedste emnesøgning i det PARADOKS-indekserede katalog.

I den første undersøgelse er antallet af poster i katalogerne, der indeholder de hyppigste emnedata, stammende fra det traditionelt indekserede katalog, sammenlignet. Der har været taget højde for denne kritik, ved at kræve at alle sammensatte emnestrengte blev søgt med nærhedsoperatører, således at ordenes rækkefølge og nærhed blev bibeholdt.

Eksempel

Emnet "Datalogi" søges på engelsk som *ordstrengen* "computer science". Derved er der ingen - eller ganske få - forvekslingsmuligheder.

Dette har i nogle få tilfælde betydet, at søgeprofilen skulle ændres, således at der ikke kom støj i resultatet. Problemerne med forvekslingsmuligheder er anført i teksten til appendiks 4. Heraf sluttes, at hvad angår det undersøgte emneområde, kan denne kvantificering - ved kraftige udsving - fortolkes som tendenser i katalogernes egnethed til emnesøgning på grove emne kategorier.

I undersøgelsen af præciseringsmuligheder tælles der ord i søgesæt, dannet i de respektive kataloger. Der er delvist taget forbehold for meningsløse ord, ved at anvende en standard stopordsliste for engelsk/amerikansk sprog og ved kun at anvende ord, der optræder i nyere abstracts fra tidsskriftsartikler om samme emne, som mål for relevans. Endelig er der ingen grund til at vurdere ordene i det traditionelle katalogs titler som værende bedre til emnepræcisering end ordene fra indholdsfortegnelser og abstracts fra det PARADOKS-indekserede katalog - tværtimod er indholdsfortegnelser og abstracts tvunget til at være indholdsbeskrivende i modsætning til bøgernes titler, der - som det er tilfældet for nærværende bog - også skal fange interessen!

Konklusion

Hovedsigtet med dette arbejde har været, at forstå nogle af årsagerne til at bibliotekernes online kataloger fungerer så dårligt til emnesøgning, at opstille et alternativ og endelig at evaluere dette alternativ.

På grundlag af analysen af emnesøgning i bibliotekernes online kataloger konkluderedes det, at problemerne nærmere ligger i selve emneindekseringen end i den måde emneindekseringerne udnyttes eller præsenteres på.

Resultatet af analysen er ikke blot, at brugerne tilsyneladende har vanskeligheder med at anvende emneordene eller den systematiske klassifikation. Resultatet af analysen er, at sådanne værktøjer - for det meste - ikke er egnede til at beskrive emnerne i vor tids videnskabelige litteratur, bundet som de er til et centralperspektiv eller en monoparadigmatisk videnskabelig kultur, der ligger nærmere middelalderens klostre end vor tids fragmenterede eller multiparadigmatiske forståelsesformer.

Årsagen hertil findes i, at det ikke er muligt at rekonstruere baggrunden eller fortolkningsrammen for indeksørens ekstremt kortfattede tolkning i form af ganske få enkeltstående ord eller klassifikationskoder.

Alternativet er at udnytte teksten fra dokumenterne systematisk til emneindeksering, idet denne tekst reflekterer den aktuelle videnskabelige kultur.

For at højne kvaliteten af emnesøgningerne i online katalogerne skal dokumentbeskrivelsen være meget fyldigere, end det er tilfældet i dag.

Men dette udløser for større bibliotekskataloger en situation, hvor alene tekstmængdens størrelse betyder, at antallet af fund, der allerede er for stort i det fleste emnesøgninger, forøges.

Så den ene forbedring af emnesøgningen fører et nyt problem med sig.

For at fastholde forbedringen uden at få de ødelæggende bivirkninger med, foreslås det, at emnesøgning i online kataloger bør foregå i en delmængde af dette katalog, bestående af oversigtslitteratur, således at emnesøgning kommer til at foregå i to trin: Først findes oversigtslitteraturen. Dernæst genbruges oversigtslitteraturens overblik over emnets litteratur.

I bogens anden del vender vi os til nogle reelt eksisterende kataloger, der anvender naturlig tekst i stedet for kontrollerede emnedata til beskrivelse af dokumenternes emner.

Gennem én udenlandsk og fem danske undersøgelser er det vist, at hvis man lægger *allerede indkøbte* samleværkers indholdsfortegnelser ind i bibliotekernes online kataloger forøges antallet af titler, der stilles til rådighed for søgning med mellem 400% og 600% for en arbejdsindsats, svarende til én procent af bogbudgettet! Med de store muligheder for emnesøgning, der ligger i samleværkernes tematisk organiserede titler, bliver samleværkerne og den i første del fremhævede oversigtslitteratur den del af litteraturen, der skal udgøre grundstammen i et alternativ til det traditionelt emneindekserede bibliotekskatalog.

Da den marginale litteratur anses for helt central for vidensudviklingen, og da denne litteratur ikke er særlig godt repræsenteret i oversigtslitteraturen, tilføjes denne litteratur, således at det er oversigtslitteraturen, samleværkerne og den marginale litteratur, som udgør det alternative emnesøgningskatalogs dokumenter.

I tråd med resultaterne fra bogens første del foretages emneindekseringen på denne litteratur alene ved hjælp af indholdsfortegnelser og abstracts, eventuelt suppleret med andre dele af dokumentet hvor indholdsfortegnelse og abstract ikke er dækkende for dokumentets indhold.

Denne indeksering, kombineret med den beskrevne dokumentudvælgelse, udgør et PARADOKS-indeksret katalog.

Herefter er der beskrevet to eksempler på PARADOKS-indekserede kataloger, det ene bestående af såvel tidsskriftartikler som bøger og det andet - benævnt det "reducerede" PARADOKS-katalog - bestående udelukkende af bøger. Begge kataloger er udviklet til at dække Danmarks Pædagogiske Biblioteks litteraturanskaffelser om "undervisning i de naturvidenskabelige og de tekniske fag" samt litteraturen om "anvendelse af undervisningsmedier", publiceret i perioden 1985-89.

Den beskrevne litteratur på Danmarks Pædagogiske Bibliotek dækker 1344 bøger, mens det fulde PARADOKS-indekserede katalog er på 560 poster og det "reducerede" PARADOKS-katalog er på 289 poster - en reduktion til henholdsvis 42% og 21% af det traditionelle katalog.

Det viser sig, at de PARADOKS-indekserede kataloger har en overvægt af engelsksproget litteratur. Det viser sig også, at den litteratur, der er udvalgt som værende marginal, hovedsagelig er publiceret som artikler i tidsskrifter, hvilket medfører, at den marginale litteratur er svagt repræsenteret i det "reducerede" PARADOKS-indekserede katalog.

Ved at sammenligne emnesøgningsmulighederne i det traditionelle bibliotekskatalog og de to PARADOKS-indekserede kataloger over oversigtslitteraturen, finder vi, at 3 ud af 4 af det traditionelle katalogs kontrollerede emnebeskrivelser umiddelbart genfindes i de PARADOKS-indekserede kataloger. Det betyder, at sådanne traditionelle søgninger i 3 ud af 4 tilfælde giver adgang til genbrug af oversigtslitteraturens beskrivelse af emnets litteratur.

Men mindst lige så interessant er det, at begge de PARADOKS-indekserede kataloger langt overgår det traditionelle katalog i præciseringsmuligheder, selv indenfor de søgesæt, der er dannet ved hjælp af det traditionelle katalogs systematiske emneklassifikation. Disse resultater er opnået på trods af, at evalueringen foregår på det traditionelle katalogs præmisser. Søgningerne er begrænset til at være søgninger på det traditionelle katalogs emneord og den systematiske klassifikations navne. Enhver søgning med synonymer eller mere specifikke emner eller metoder, der ikke er indeholdt i

de kontrollerede sprog, vil give nye muligheder for genbrug af oversigtslitteraturen.

Endelig skal det understreges, at den foretagne evaluering kun den del af de PARADOKS-indekserede kataloger, der dækker oversigtslitteraturen. Den marginale litteraturs bidrag til de PARADOKS-indekserede katalogers resultater ikke er medtaget eksplicit.

Umiddelbart ser det PARADOKS-indekserede katalog ud til at være et fristende alternativ til det traditionelle online katalogs emneindekseringer:

- det har ca. 75% af det traditionelle katalogs emnesøgningsmuligheder indbygget
- det har præciseringsmuligheder, der i langt de fleste tilfælde overgår det traditionelle katalogs
- det øger antallet af søgbare titler med 400-600% uden at der skal købes én eneste bog mere
- det præsenterer overkommelige søgesæt, tematisk organiseret af en navngiven person i stedet for de alt for store perspektivløst samlede søgeresultater, der opstår ved emnesøgning i de traditionelle online kataloger
- det anvender emneområdets eget sprogbrug og følger automatisk med den vidensudvikling, der foregår i fagets publikationer
- det anvender kun kontrollerede sprog til brug for ganske få overordnede kategorier
- det koster mindre at producere end emneindekseringen i det traditionelle onlinekatalog

Men der skal foretages yderligere undersøgelser for at kunne generalisere fra denne undersøgelses resultater.

For at kunne udtale sig alment om det PARADOKS-indekserede katalogs dækningsgrad i forhold til det traditionelt emneindekserede katalog, skal der foretages undersøgelser på andre fagområder med andre - mere flertydige - begreber, end det er tilfældet for de undersøgte områder.

Og det skal vurderes, om den yderligere betoning af engelsksproget litteratur, der følger af udvælgelsen af oversigtslitteratur og samleværker til emnesøgning, også er acceptabel for "lokale" og "små" emneområder.

English summary

A steadily increasing amount of scientific publications, combined with a fragmentation of the scientific cultures, raises great difficulties in the use of library catalogues for subject access to literature. Global qualitatively based selection procedures are inhibited by the huge amount of publications. And the standard tools for subject access - classification and key words - are inappropriate in a highly fragmented scientific culture.

These observations are - at the dawn of the information age - alarming signs of a break down of a public information service, leaving the scene to commercial and informal networks, out of public control.

THE FIRST PART OF THIS BOOK consist of a description of the theoretical and empirical basis for a new way to build and organise subject access in library catalogues.

In the light of the fact that library catalogues only forms a constituent part of the means for subject access, we examine if library catalogues are used for subject access and what could be done to improve this facility. It seems, that substantial improvements includes the use of so-called natural texts - e.g. tables of contents or abstracts - as subject description of documents, represented in online catalogues.

Such an improvement has, on the other hand, often a very serious drawback as the amount of documents found will explode without any available qualitatively selection criteria.

One possible solution is to give up the tradition, trying to adapt the whole library catalogue for subject access. Subject access could be restricted to be executed in the survey literature, using the natural indices - table of contents or abstracts - as subject description.

This raises the old controversy between the advocates for subject indexing in controlled languages and the advocates for the use of natural text. The advantages and disadvantages, using the two methods for subject description, are resumed, and the problem of indexing consistency is reviewed.

Some aspects of a theory of subject indexing and subject access is introduced using the concept of a model in classical logic and the concept of band width in information theory. Both considerations stresses the need for authentic subject descriptions in library catalogues, quite different from the dominating praxis, using classmarks and key words, assigned by an indexer, obscure to the user.

THE SECOND PART OF THE BOOK introduces practical investigations, build upon the results in the first part.

A new concept of online catalogues - called PARADOKS-indexed catalogues - is developed.

The subject access process, using a PARADOKS-indexed catalogue, is split in two processes.

First process is a search exclusively in the survey literature and the marginal literature. The searches are exclusively executed as word-searches in the "natural indices" of the documents as the tables of contents and abstracts. No keywords or classmarks are added.

The second process of the search consist of reading the literature and subject overview of the survey literature. This lead to a high quality paradigm specific subject access to the literature, described in the surveys.

Survey literature appears in different forms. Bibliographies, state of the arts, literature reviews and composite works.

Composite works constitutes a substantial part of the survey literature, giving the editors selection of articles on a given subject. Quantitatively, composite works is the dominating document typology in the survey literature. Therefore we examine the order of magnitude of the extension of survey literature in library catalogues by analysing the amount of composite works in library catalogues. It is shown by 5 random samplings in 2 danish scientific libraries, that composite works - in average - amount to 20% of the collections. It is also shown, that each volume - in average - is composed by 20 articles. These results are comparable to an analysis, made on an american college library. This means that libraries only shows around 20% of their titles in their catalogues, not analyzing the constituent articles in the composite works.

Finally a PARADOKS-indexed catalogue is described and evaluated. The PARADOKS-indexed catalogue is constructed as an alternative subject access to 1344 documents in an traditional catalogue.

It is shown, that more than 75% of the 194 most frequent keywords and classmarks in the traditional catalogue already appears in the free text of the tables of contents and abstracts in the PARADOKS-indexed catalogue.

Still - the traditional catalogue could give superior subject access if the further possibilities for further relevant subject information exceeded the possibilities in the PARADOKS-indexed catalogue. Comparing the results from searches on the 25 most frequently classmarks shows that in more than 80% of the cases, the PARADOKS-indexed catalogue improved the possibilities for relevant subject precision compared to the traditional library catalogue.

APPENDIKS 1

51 EMNEOMRÅDER MED ENCYKLOPÆDISKE ARTIKLER

Med reference til omtalen på side 20 bringes her en liste med 51 emneområder med angivelse af tilhørende encyklopædiske artikler fra Lancasters undersøgelse^a.

^a Lancaster, Connell, Bishop & McCowan 1991.

FIGUR 18
51 emner med tilhørende 51 encyklopædiske artikler fra Lancaster et.al. 1991.

FIGUR 19
51 emner med tilhørende 51 encyklopædiske artikler fra Lancaster et.al. 1991.
Fortsat.

OM STATISTIKKEN

I det følgende opridses det statistiske grundlag for stikprøveanalyserne i kapitel 5.

Som gennemgående eksempel anvendes den anførte måling på ca. 200.000 monografier på Danmarks Pædagogiske Bibliotek, publiceret fra lige før århundredeskiftet til 1990.

Bøgerne er systematisk opstillet, og ved at udtage sidste bind på 4' og 5' hylde i hver reol sikrer vi os, at stikprøven ikke udvælger bestemte dokumenttyper.

Vi måler antallet af værker i hver enkelt dokument i stikprøven, og stikprøven udtager 718 dokumenter.

ANDELEN AF SAMLEVÆRKER

Ud af de $N=718$ dokumenter var $N_s=114$ samleværker, det vil sige dokumenterne indeholdt 2 eller flere værker.

Stikprøven viste sig altså at indeholde

$$x = \frac{N_s}{N} 100\% = 15,88\% \text{ samleværker.}$$

En fornyet stikprøve vil selvfølgelig med stor sandsynlighed have en anden andel af samleværker, men denne variation er begrænset af, at en vilkårlig ny prøve med 95% sandsynlighed vil ligge i et

interval omkring den fundne værdi på det dobbelte af spredning på $x \pm 2\sigma(x)$, der er på 1,36%.

Det vil sige at vi med 95% sikkerhed kan sige at antallet af samleværker i de ca. 200.000 bind ligger i intervallet $16 \pm 3\%$.

Årsagen til at vi kan udtale os om dette er skyldes to forhold. For det første er stikprøven binominalt fordelt, hvilket betyder at spredningen på fordelingen x er

$$\sigma(x) = \sqrt{\frac{x(1-x)}{N}}$$

For det andet er såvel antallet af dokumenter xN og $(1-x)N$ i stikprøvens kategorier tilstrækkelig store til at fordelingen nærmer sig en normal fordeling, hvor netop 95% af stikprøverne vil ligge indenfor et interval på to gange spredningen omkring gennemsnittet fra den første stikprøve.

2. Antallet af værker i samleværkerne.

Ved hjælp af stikprøven kan vi også udtale os om antallet af bidrag fra samleværkerne.

I denne måling udtages 114 stikprøver, hvor antallet af bidrag, r_i , tælles. Idet s angiver at der er tale om et samleværk, får vi en fordeling af et antal samleværker $N_{s,r}$, hvor $r=2,3,4,5,\dots$ angiver antallet af bidrag fra hvert samleværk.

Hver af disse fordelinger er - som før - binominale fordelinger. Fordelingerne er uafhængige af hinanden og har hver deres middelværdi og spredning.

For at finde det gennemsnitlige antal bidrag pr. samleværk, middelværdien Y , summeres disse fordelinger og divideres med den samlede mængde af samleværker i stikprøven:

$$Y = \sum \frac{rN_{s,r}}{N_s}, \text{ hvor } N_s = \sum_{r=2}^{\infty} N_{s,r}$$

I den aktuelle stikprøve - se fig. Figur 9 - er middelværdien $Y=20,04$.

For de værdier af r , der er tale om her, gælder det - ifølge den såkaldte centrale grænseværdisætning - at denne sum er praktisk taget normalt fordelt. Og når middelværdien Y selv er praktisk taget normalt fordelt kan vi udlede, at middelværdien Y ved gentagelse af stikprøven med 95% sikkerhed vil ligge indenfor et interval på to gange spredningen på *middelværdien* omkring den første værdi af Y . Spredningen på middelværdien Y kan vises at være^a

$$\sigma(Y) = \sqrt{\frac{\sum_{r=2}^{\infty} (rN_{s,r} - Y)^2}{N_s(N_s - 1)}}$$

I den aktuelle stikprøve er $\sigma(Y)=2,12$.

Det vil sige, at vi med 95% sikkerhed er mellem 16 og 24 værker pr samleværk.

Ovenstående betragtninger over statistikken er udelukkende eksemplificeret ved hjælp den ene af de tre udførte analyser, men gælder naturligvis for dem alle.

^a Arley 1961:§11.7.

PARADOKS-INDEKSEREDE POSTER

Nedenfor er der givet nogle eksempler på indholdet af poster i den PARADOKS-indekserede base på Danmarks Pædagogiske Bibliotek.

SAMLEVÆRK

Her beskriver indholdsfortegnelsen dokumentets emne tilstrækkeligt.

Schoenfeld, Alan H. (ed.): Cognitive Science and Mathematics Education. Lawrence Erlbaum Associates, Hillsdale, N.J. 1987.

INDHOLDSFORTEGNELSE

1. Cognitive Science and Mathematics Education: An Over-view. Alan H. Schoenfeld.
2. Foundations of Cognitive Theory and Research for Mathematics Problem-Solving. Edward A. Silver.
3. Instructional Representations Based on Research about Understanding. James G. Greeno.
4. Cognitive Technologies for Mathematics Education. Roy D. Pea.
5. Problem Formulating: Where Do Good Problems Come From? Jeremy Kilpatrick.
6. From the Teacher's Side of the Desk. Anna Henderson.
7. New Knowledge about Errors and New Views about Learners: What They Mean to Educators and More Educators Would Like to Know. Stephen B. Maurer.
8. What's All the Fuss about Metacognition? Alan H. Schoenfeld.

9. Cognitive Science and Algebra Learning. Ronald H. Wenger.
10. Cognitive Science and Mathematics Education: A Mathematician's Perspective. H. O. Pollak.
11. Cognitive Science and Mathematics Education: A Mathematics Educator's Perspective. F. Joe Crosswhite.

MONOGRAFI

Her suppleres indholdsfortegnelsen med et uddrag fra selve teksten for at emnebeskrivelsen er tilstrækkeligt dækkende.

Technological Revolution. The Politics of School Science & Technology in England & Wales since 1945./ Gary McCulloch, Edgar Jenkins and David Layton.
The Falmer Press 1985.

INDHOLDSFORTEGNELSE

1

Two cultures? Guardians of Science. The Alternative Road. Science and Politics. Royal Patron of Engineering. The Nuffield Foundation Science Teaching Project. Engineering among the schools. Projects in Technology. The Schools Science and Technology Committee. The Great Debate.

UDDRAG FRA SELVE TEKSTEN

1

The book examines the role of science teaching associations and the post-war secondary technical schools in curriculum change, and assesses the importance not only of such agencies as the DES, the Schools Council and the Manpower Services Commission but also interest groups related to industry, engineering and professional science. The changing situation since the opening of the Great Debate on Education in 1976 is described in detail. The book also considers the impact and significance of recent curriculum initiatives and examines the prospect for future reform.

TIDSSKRIFTSARTIKEL

Her anvendes ord fra artiklens abstract, kombineret med uddrag fra teksten, bestående af mellemrubrikker m.v..

Streibel, Michael J.: "A Critical Analysis of the Use of Computers in Education" ECTJ (Educational Communication and Technology) 34(1986)3,137-161.

ABSTRACT, FORORD, M.V.

1

technologizing education, sub-skill acquisition, delegitimize non-technological ways of learning and thinking about problems

UDDRAG FRA SELVE TEKSTEN

1

tacit knowledge

2

ambiguous knowledge

3

Problematic knowledge

4

drill and practice computer programs, mastery learning, individualization, courseware as replicated work

5

tutorial computer programs, human computer dialogues, humans are treated as data based rule-following symbol - manipulating information processors, education will be viewed as a form of training and will be subject to explicit extensional logic

6

computers as intellectual tools, problem-solving tool, computer programming and computer simulations

EMNEORD I PARADOKS-BASEN

I Danmarks Pædagogiske Biblioteks katalog blev de 1344 poster, produceret i perioden 1985 til 1989, over litteratur, publiceret i samme periode og indkøbt og opstillet indenfor emneområderne undervisningsmedier, naturvidenskabernes og de tekniske fags pædagogik analyseret for anvendelse af emneord.

Ingen poster, inddateret før 13/11 1990 er medtaget.

Ved anvendelse af den indbyggede statistik funktion i Danmarks Pædagogiske Biblioteks katalog blev der produceret en liste af emneordsstreng, der var brugt 4 gange eller mere.

Det drejer sig om 157 emneordsstreng.

Emneordene blev direkte oversat ord til ord til tysk og engelsk, der indgik i søgeprofilerne.

Enkelte ord er ikke afsøgt i alle felter, nemlig de ord, der vil kunne give anledning til oplagt støj. Et eksempel er "Universitet", der - specielt på engelsk - indgår i mange forlagsnavne.

Der er udelukkende anvendt nærhedsoperatorer, når der søgt med søgestreng. Således er f.eks. "computer science" søgt som "computer adj science", hvor adj angiver, at ordene skal følge hinanden i den angivne rækkefølge.

Ved hjælp af disse søgeprofiler blev der foretaget søgninger i det PARADOKS-indekserede katalog.

Emneord	Fund, DPB	Søgeprofil, PARADOKS	Fund, PARADOKS	
			total	bøger
danmark	352	d?n\$I mark	17	13
storbritannien	157	england or britain	20	18
forsøg	152	forsøg or experimentation? Or schulers?Ch\$	6	5
fag i læseplan	149	curriculum adj subject or lehrfach or fag near2 læseplan or skolefag\$	1	1
sverige	133	sverige or sweden or schveden	17	12
klasetrin syv til ti	132	overbygning\$ or lower adj secondary or mittelstufe\$ or melemtrin\$	4	4
tyskland forbundsrepublik	96	tyskland or vesttyskland or germany or brd or deutschland	15	11
datalogi	93	dataal\$ or informatic\$ or informatik\$ or computer adj science	44	24
usa	90	usa	13	10
uddannelse på primaert niveau	85	begynderundervisning\$ or primary adj education or primarbildung	4	4
konsroller	75	konsrolle\$ or (sex or gender) with role\$ or geschlechtsrolle\$	5	4
forsøg folkeskole	69	forsøg\$ with folkeskole\$ or experimentation? With (primary or secondary) or schulers?Ch\$ with (primarbild\$ or mittelstufe?)	1	1
dataformidlet undervisning	66	Dataformidlet or computer adj assisted or computerunterstütz\$	49	35

Emneord	Fund, DPB	Søgeprofil, PARADOKS	Fund, PARADOKS	
			total	bøger
informationsteknologi informations- samfund og undervisning	57	(information\$te\$log\$ or information adj te\$log\$) with (informationssamfund\$ or information adj societ\$ or informationsgesellschaft\$)	1	1
matematik	56	mat\$temati\$	159	103
voksenuddannelse	51	voksenuddannelse\$ or adult adj education or erwachsenenbild\$	9	7
prøveforb voksenundervisning	50	prøveforberedende adj voksenundervisning or continuation adj education or zweiter adj bildungs- weg	0	0
computer	48	computer	168	104
kvinders uddannelse	48	kvinde\$ with \$uddannelse or wom?N\$I with education or frauenbildung\$	7	6
voksenundervisning kompetencegi- vende	47	(erwachsenen\$ or voksen\$ or adult\$) with (kompeten\$ or qualif\$)	0	0
datakøre	46	data\$ or informatic\$ or informatik\$ or (computer adj science)	44	24
voksenundervisning forsøgs- og udviklingsarbejder	46	(erwachs\$ or voksen\$ or adult\$) with (forsøg\$ or udviklingsarbejde\$ or experimentation\$ or versuch\$)	0	0
10 punktsprogrammet	44	10 adj punktsprogrammet	0	0
norge	43	norge or norway or norwegen	6	5
integreret læseplan	37	integr\$ same (læseplan or bildungsplan or curricul\$)	18	13

Emneord	Fund, DPB	Søgeprofil, PARADOKS	Fund, PARADOKS	
			total	bøger
klassetrin elleve til tretten	37	(gymnasie\$ or htx or erhvervsfaglig\$ adj grunduddannelse\$ or upper adj secondary or oberstufe	8	7
teknologi	37	teknologi\$ or technolog\$	197	119
forsøg ungdomsuddannelser	33	(experimentation\$ or versuch\$ or forsøg\$) same (ungdomsuddannelse\$ or post adj obligatorisches or post adj compulsory)	0	0
undersøgelse	32	undersøgelse\$ or erhebung\$ or survey\$[-udvk]	30	22
miljølere	29	miljølere or umweltforschung\$ or environmental adj stud\$	3	2
naturvidenskaber	29	natural adj science\$ or naturwissenschaft\$ or naturvidenskab\$	25	24
undervisningsplaner	24	undervisningsplan\$ or lehrplan\$ or curricul\$	119	84
højere uddannelse	23	højere adj uddannelse or higher adj education or hochschulbildung	22	15
kunstig intelligens	23	kunstig adj intelligens\$ or artificial adj intelligens\$ or ai or künstlich\$ adj intelligens\$	35	27
officiel tekst	23	offi?iel adj (tekst\$ or text\$)	0	0
erhvervsmæssig grunduddannels	22	erhvervsmæssig adj grunduddannelse\$ or berufsgrundbildung\$ or basic adj training	0	0
fysik	22	\$fysik\$ or \$physics\$ or physics	50	39
temaundervisning	22	temaundervisning or thema\$	116	7

Emneord	Fund, DPB	Søgeprofil, PARADOKS	Fund, PARADOKS	
			total	bøger
edb program	21	(edb or computer) adj program\$	7	6
skolepræstationer	21	\$præstationer\$ or (achievement\$ or leistung\$)	36	24
afgangsprøve	20	afgangsprøve\$ or final adj examination\$ or abschlussprüfung\$	1	1
kønsforskel	20	kønsforskel\$ or sex\$ same \$differen\$ or geschlechtsunterschied\$	9	5
spørgeskema	20	spørgeskema\$ or questionnaire\$ or fragebog\$	6	4
regning	19	regning or calcul\$ or rechner\$	25	18
teknisk uddannelse	19	teknisk same uddannelse\$1 or technical adj education or technische\$ adj bildung\$	7	7
bibliografi	16	bibliograph\$	23	15
biologi	16	biolog\$	26	18
projektmetode	16	projekte\$ or project adj method	0	0
holland	15	holland or netherlands\$ or niederl\$	24	21
kemi	15	kemi\$ or chemi\$	27	19
kvinduddannelse	15	kvindes\$ same \$uddannelse\$ or wom?N\$ same education or frauenbildung\$	13	10
fjernundervisning	14	fjernundervisning or distance with education or fernunterricht\$	20	15

Emneord	Fund, DPB	Søgeprofil, PARADOKS	Fund, PARADOKS	
			total	bøger
værksted	14	værksted\$ or workshop\$ or werstatt\$	20	19
erhvervsuddannelse	13	erhvervsuddannels\$ or vocational adj education or berufsbiildung\$	9	8
europa	13	europ\$	20	17
information	13	information	106	65
konference	13	?Onference\$[-dt]	45	37
videre uddannelse af lærere	13	videre with uddannelse\$ same lærer\$ or further adj education with teacher\$ or lehrfortbiildung\$	0	0
curriculum	12	curricul\$	118	84
ernæring	12	ernæring\$ or nutritiion\$ or ernährung\$	5	3
ingeniør	12	\$ingeni\$ or engineer\$	38	34
kvinde	12	kvinde\$ or wom?N\$ or frau\$	47	36
læreruddannelse	12	læreruddannelse\$ or teacher with education or lehrerbildung	36	28
simulation og gaming	12	simulation\$ or gaming	45	31
tekniske videnskaber	12	teknisk\$ with videnskab\$ or engin\$ adj science\$ or ingenieurwissenschaften\$	1	0

Emneord	Fund, DPB	Søgeprofil, PARADOKS	Fund, PARADOKS	
			total	bøger
tyskland ddr	12	østtyskland or ddr or deutsche adj dr or german adj dr	1	1
fremtidsforskning	11	fremtidsforskning\$ or futurolog\$	1	1
husholdningslære	11	husholdningslære\$ or home adj economics or hauswirtschaft\$	1	1
skriftligt arbejde	11	skriftlig\$ adj arbejde\$ or written adj work or schriftlich\$ adj arbeit\$	0	0
indskoling	10	indskoling\$ or entr\$ adj2 school or schuleintritt\$	1	0
kvalifikation	10	kvalifikation\$ or qualifi?ation\$	9	8
massekommunikation	10	masse\$kommunikation\$ or mass adj communication	0	0
norden	10	norden or s?Andinavi\$	3	3
udflugt	10	udflugt\$ or excursion\$ or ausflug\$	0	0
arbejdsmarked	9	arbejdsmarked or labour adj market or arbeidsmarkt	1	1
efter og videreuddannelse	9	efteruddannelse\$ or videreuddannelse\$ or further adj eucation or weiterbildung\$	4	3
evaluering	9	evalu\$	80	52
finland	9	fin\$lland	10	9
informationsøgning	9	informationsøgning or information adj retrieval	5	5

Emneord	Fund, DPB	Søgeprofil, PARADOKS	Fund, PARADOKS	
			total	bøger
informationsteknologi	8	information\$1e\$2nolog\$ or information adj technolog\$	53	35
skolesystem	8	skolesystem or school adj system\$ or schulsystem\$	6	5
universitet	8	universit\$[-bi]	56	43
zimbabwe	8	zimbabwe	2	1
arbejds kvalifikation	7	arbejds kvalifikation\$ or berufsant?Rderung\$ or job same requirement\$	0	0
database	7	dat\$1bank\$ or data adj bank\$	1	1
datateknik	7	datateknik or computertechni\$ or computer adj engineering	2	2
elektronik	7	ele?Tromi?\$1	24	20
erfaring	7	Erfaring\$ or experience\$ or erfahrung\$	86	59
lokalsamfund	7	lokalsamfund\$ or gemeind\$ or local adj communiti\$	1	1
pige	7	pige\$ or girl\$ or mädchen\$	38	28
pædagogisk filosofi	7	pædagogisk\$[-op] adj filosofi\$ or erziehungphilosophi\$ or philosophi\$ near2 education	1	1
uddannelsessystem	7	uddannelsessystem\$ or educational near2 system\$ or bildungssystem\$	16	13

Emneord	Fund, DPB	Søgeprofil, PARADOKS	Fund, PARADOKS	
			total	bøger
dansk sprog	6	dansk adj sprog or danish adj language\$ or danske adj sprach\$	0	0
england	6	england or britain	20	18
erhvervsforberedende uddannelser	6	erhvervsforberedende adj uddannelse\$ or (prevocational or pre adj vocational) adj education or berufsvorbereitende adj ausbildung\$	0	0
futurologi	6	futureolog\$	1	1
geografi	6	geogra\$	17	12
holdning	6	holding\$ or einstellung\$ or attitude\$	58	36
immigrant og minoritetsforskning	6	immigrant\$ or minorit\$	7	6
læreprocesser	6	læreprocesser\$ or lernprozess\$ or learning adj process\$	17	12
modulstruktureret oplæring	6	modul\$	16	15
samfund i	6	samfund\$ or gesellschaft\$ or societ\$	65	50
uddannelsesinstitution	6	uddannelsesinstitution\$ or bildungseinrichtung\$ or educational adj institution\$	2	1
uddannelses kvalitet	6	uddannelseskvalitet\$ or bildungsqualit\$ or quality near2 education	1	1
uddannelsesplanlægning	6	uddannelsesplanlegung\$ or planning near2 education\$	3	1

Emneord	Fund, DPB	Søgeprofil, PARADOKS	Fund, PARADOKS	
			total	bøger
uddannelsesvejledning	6	uddannelsesvejledning\$ or bildungsberatung\$ or educational near2 guidance\$	1	1
universitetsstudier	6	universitetsstudier\$ or hochschulstudii\$ or universit\$ near2 stud\$	9	8
adgang til uddannelse	5	adgang with uddannelse\$ or bildungszugang\$ or access\$ with education\$	2	2
elektroteknik	5	elektroteknik\$ or electrical adj engineering	1	1
elev	5	elev\$ or schüler\$ or pupil\$	68	53
fornælse	5	fornælse\$ or innovation\$	28	18
frankrig	5	frankrig or france or frankreich	22	16
industri	5	industri\$2	16	12
intern virksomhedsuddannelse	5	intern adj virksomhedsuddannelse or (job or industr\$) with training or betriebs\$ with \$bildung\$	3	3
kognitiv udvikling	5	?Ognitiv\$ adj (udvikling\$ or development\$ or entwickl\$)	9	7
læreres efteruddannelse	5	lærer\$ near2 efteruddannelse\$ or lehrerfortbildung\$ or further near2 education near2 teachers	3	2
manuelt arbejde	5	manuel\$ near2 arbejds\$ or manual near2 work\$ or handarbeit\$	0	0
massemedier	5	massemedier\$ or mass adj media	8	6

Emneord	Fund, DPB	Søgeprofil, PARADOKS	Fund, PARADOKS	
			total	bøger
skandinavien	5		0	0
teknologisk forandring	5	s?Andnav\$ te?\$INolog\$ near2 (forandring or wandel\$ or change\$)	12	9
tilvalgsfag	5	tilvalgsfag\$ or wahlfach\$ or option\$	8	5
tyskland brd	5	tyskland or brd or deutschland or germany	15	11
uddannelse på sekundært niveau	5	uddannelse with sekundært adj niveau or sekundarbildung or secondary same education	34	26
uddannelsespolitik	5	uddannelsespolitik or bildungspoliti\$ or edcational near2 polic\$	2	1
videobåndoptagelse	5	videobåndoptagelse\$ or magnetisches bildaufzeichnung\$ or video adj recording\$ or videorecor- ding\$	1	1
voksenundervisning grundlæggende	5	voksenundervisning\$ near grundlæggende or basic adj adult or grundlegend\$ near2 erwachsen- bildung\$	0	0
voksenundervisnings metodik	5	voksenundervisning\$ same metod\$ or (adult near2 education) with method\$ or erwachsen- bildung\$ with method\$	0	0
arbejdskraftbehov	4	arbejdskraftbehov or arbejdskraftbedarff\$ or manpower near2 need\$	0	0
audio visuelle undervisningsmidler	4	(audio\$ or visuel\$) near2 undervisningsmid\$ or audiovisuel\$ near2 lehrmittel\$ or audiovisual aids	12	9

Emneord	Fund, DPB	Søgeprofil, PARADOKS	Fund, PARADOKS	
			total	bøger
curriculumudvikling	4	curriculumudvikling or curriculumentwicklung or curriculum adj development\$	19	15
database	4	database\$	22	19
elektronisk databehandling	4	elektronis\$ adj dat\$ or electronic adj data adj processing	0	0
filosofi	4	(? \$! Iloso\$)	29	16
forenede kongerige	4	forenede adj kongerige or britain or england	20	18
håndarbejde	4	håndarbejd\$ or needle\$ or handarbejt\$	2	1
informationsbehandling	4	informationsbehandling\$ or informationsverarbeitung\$ or information near2 process\$	18	9
informationssystem	4	informationssystem or information near2 system\$	9	8
klassetrin et til seks	4	klassetrin with et with seks or primarschul\$ or primary adj education	4	4
kognitiv proces	4	?Ognitiv\$ adj pro?Es\$	8	6
kommunestyre	4	kommunestyre or kreisverwaltung\$ or local adj government\$	0	0
lære	4	lære or læreling\$ or Lehre or apprenticeship\$	2	2
lærebog	4	læreb\$ or lehrbuch or textbook\$	21	17
læreproces	4	læreproces\$ or lernprozess\$ or learning adj process\$	17	12

Emneord	Fund, DPB	Søgeprofil, PARADOKS	Fund, PARADOKS	
			total	bøger
organisation	4		33	30
projekt metoden	4	organi?Ation	0	0
raceproblemer	4	projec?TS adj met\$od\$ or projektmethod\$	0	0
schweiz	4	raceproblem\$ or rassenproblem\$ or race\$ near3 problem\$	3	2
skolebespisning	4	schweiz or switzerland	0	0
statistisk analyse	4	skolebespisning or schulspeisung or school with meals	3	2
tekstbehandling	4	statistis\$ with analys\$	13	11
telekommunikation	4	tekstbehandling or word adj proces\$	15	12
tests	4	tele?Ommuni?Ation\$ or fermeldewesen\$	36	31
turisme	4	test\$3	0	0
uddannelsesbehov	4	turism\$ or tourism\$	1	1
uddannelsesmuligheder	4	uddannelsesbehov\$ or bildungsbed\$ or educational adj need\$	1	1
		uddannelsesmuligheder\$ or bildungschance\$ or educational adj opportunit\$		

SYSTEMATISKE KLASSIFIKATIONS- KODER I PARADOKS-BASEN

I Danmarks Pædagogiske Biblioteks katalog blev de 1344 poster, produceret i perioden 1985 til 1989, over litteratur, publiceret i samme periode og indkøbt og opstillet indenfor emneområderne undervisningsmedier, naturvidenskabernes og de tekniske fags pædagogik analyseret for anvendelse af klassifikationskoder. Ingen poster, inddateret før 13/11 1990 er medtaget.

Ved anvendelse af den indbyggede statistik funktion i Danmarks Pædagogiske Biblioteks katalog blev der produceret en liste af klassifikationskoder, der var brugt 4 gange eller mere.

Det drejer sig om 81 klassifikationskoder. Af disse var 37 indenfor undersøgelsens emneområde.

Klassifikationskodernes navne eller sproglige ækvivalenter blev direkte oversat ord til ord til tysk og engelsk, der indgik i søgeprofilerne.

Der er udelukkende anvendt nærhedsoperatorer, når der søgt med søgestrengene. Således er f.eks. "Computer assisted" søgt som "computer adj assisted", hvor adj angiver, at ordene skal følge hinanden i den angivne rækkefølge.

Ved hjælp af disse søgeprofiler blev der foretaget søgninger i det PARADOKS-indekserede katalog.

Klass.	Fund, DPB	Forklaring	Søgeprofil, PARADOKS	Fund, PARADOKS	
				Total	Bøger
m51	248	Matematik. Metodik.	Matematik\$ or matematik\$	156	100
37.1322	224	Datamaskinformidlet undervisning.	Datamaskinformidlet or computer adj assisted or computerunterstit\$	48	34
m51.1	222	Datalogi. Informatik. Metodiske bøger	data\$ or informatic\$ or informatik\$ or (computer adj science)	44	24
m50	173	Naturkundskab. Metodik.	Natur\$ adj science\$	13	10
m60	152	Anvendt videnskab, teknik, praktiske fag i almindelighed. Metodik.	Anvendt\$ adj videnskab\$ or applied adj science\$ or angew\$ adj wis-senschaft\$	0	0
37.139	84	Mediepædagogik og undervisningsmidler i alm.	Mediepædagog\$ or education\$ near2 media\$ or medienpædagog\$	27	11
m50.1	60	Naturen og mennesket. Forurening. Økologi. Metodiske bøger	forening\$ or ecolog\$ or økolog\$ or verschmutz\$ or økolog\$	11	9
37.155	60	Evalvering af undervisningsmaterialer, herunder audio-visuelle materialer, datamaskinformidlede programmer m.v.	(\$Undervisningsmaterial\$ or educational adj material? Or teaching adj aids or lehrmittel\$) same (\$analyse\$ or research\$ or \$forschung\$ or evaluat\$)	0	0
37.132	48	Audio-visuelle hjælpemidler, litteratur om, herunder film, radio og fjersyn som undervisningsmiddel.	(Audio\$ with visuel\$) or av or radio? Or \$fjernsyn\$ or television\$ or tv	50	29

Klass.	Fund, DPB	Forklaring	Søgeprofil, PARADOKS	Fund, PARADOKS	
				Total	Bøger
m53	47	Fysik. Metodik.	Physics or \$fysik\$ or \$physik\$	50	39
m58	42	Biologi som undervisningsfag. Metodik.	(\$Biolog\$ or botan\$ or \$zoolog\$)	29	20
m40	34	Undervisning i geografi.	\$Geogra\$	18	13
51.18	33	Datamaskiner og informatikinstitutioner. Om maskinpark, softwareproduktion og distribution, jura samt institutioner.	Kan ikke oversættes	0	0
m54	31	Kemi. Metodik.	Kemi\$ or chemi\$	27	19
51.7	27	Matematikvanskeligheder.	(Matematik\$ or mathemati\$ or calculus\$ or regne\$) same (\$vanskelighed\$ or difficult\$ or \$störung\$) or dys?AI\$	8	7
64.3	26	Madlavning. Ernæringslære. Servering. Borddækning.	Madlavning\$ or ernæring\$ or cooking or nutrition or kochen or ernährung\$	5	3
51.9	23	Undervisningsfaget "matematik og samfund", herunder undervisning i matematikken historisk. Undervisningsfaget "informatik og samfund", herunder undervisning i informatikkens historie.	(\$Matematik\$ or mathemati\$) with (samfund\$ or societ\$ or gesellschaft\$ or ?\$Hlloso\$ or histor\$)	19	13

Klass.	Fund, DPB	Forklaring	Søgeprofil, PARADOKS	Fund, PARADOKS	
				Total	Bøger
65.8	19	Færdselslære. Trafikundervisning.	\$Færdsel\$ or traffic\$ or \$trafik\$	4	4
60.7	14	Teknisk skole. Håndværkerskoler. Specialarbejderskoler. Tekniske højskoler. Institutionsbeskrivelser.	(Teknik\$ or technolog\$ or techn\$ or ingen\$ or engine\$) near2 (institution\$ or school\$ or \$skole\$ or \$schul\$)	21	17
60.9	12	Undervisningsfaget "teknik og samfund". Historisk og samfundsmæssig sammenhæng.	(\$Teknik\$ or technolog\$ or \$techn\$ or \$ingen\$ or engine\$) same (samfund\$ or societ\$[-bi] or gesellschaft\$ or ?\$[I]llos\$)	33	21
62.01	10	Energiteknik og -politik.	Energy\$ same (technolog\$ or \$teknologi\$ or \$politik\$ or polic\$)	2	2
68.5	10	Snedkerhåndværk, træsløjd.	Handcraft\$ or \$sløjd\$ or woodwork\$ or werken\$	1	1
62.3	9	Elektroteknik, elektronik.	Ele/Trom\$ or ele/Trote\$	29	23
64.09	9	Husvæsenundervisnings institutionelle rammer og deres historie.	(Husholdning\$ or home adj economic\$ or haushalt\$) near2 (institution\$ or \$kol\$ or schul\$ or school\$ or history)	0	0
64	9	Husvæsen. Husholdning. Hjemkundskab.	Husholding\$ or home adj economic\$ or haushalt\$	4	3
64.6	9	Tekstiltilvirkning, ikke-industriel. Håndarbejde. Her anbringes også metodiske bøger	håndarbejde\$ or needlework\$ or handarbeit\$ or textil\$	3	3

Klass.	Fund, DPB	Forklaring	Søgeprofil, PARADOKS	Fund, PARADOKS	
				Total	Bøger
69	8	Byggeteknik. Arkitektur.	Byggeteknik or arkitekt\$ same uddannelse\$ or architect\$ same education\$	3	2
62	8	Maskinlære, energiteknik, varmeteknik, elektronik i almindelighed.	Maskinlære or energiteknik\$ or varmeteknik\$ or mechanical adj engineer\$ or energy adj technology\$ or heat adj engineer\$ or maschin\$bau\$ or energietechnology\$ or wärmetechnology\$	1	0
50.9	8	Undervisningsfaget "naturvidenskab og samfund", herunder undervisning i naturvidenskabens historie.	Natur\$ same science\$ same (samfund\$ or societ\$ or gesellschaft\$ or ?\$Illo- so\$)	9	4
66.8	8	Levnedsmiddelindustri. Industriel madlavning.	Food\$ same (technology\$ or industr\$) or levnesmid\$ or lebensmittel\$	0	0
m63	6	Landbrug, skovbrug, gartneri, fiskeri, jagt. Metodik.	Landbrug\$ or farming\$ or agricultur\$ or landwirtschaft\$	4	2
60.2	5	Materiallære i almindelighed, materialeprøvnings i almindelighed. Produktkontrol i almindelighed.	Materiallære or material\$ adj science or materialkunde	1	0
64.7	5	Husvesenets økonomi samt kollektivets og institutioners husholdning.	(Husholdning\$ or (home adj econom\$) or haushalt) same (institution\$ or school\$ or \$schule\$ or \$sko\$)	1	0

Klass.	Fund, DPB	Forklaring	Søgeprofil, PARADOKS	Fund, PARADOKS	
				Total	Bøger
m55	4	Geologi. Metodik	\$geolog\$	10	4
62.015	4	Vedvarende energikilder, alternative energikilder og teknikken til disses udnyttelse.	(Alternativ\$ or vedvarende or renewable or dauern\$) adj energ\$	0	0
35.61	4	Skolefjernsyn. Fjernsyn som undervisningsmedier.	Skolefjernsyn\$ or school adj television\$ or schulfjernseh\$	0	0
63.07	4	Landbrugsundervisning. Undervisning i skovbrug, gartneri, fiskeri og jagt. Institutionelle beskrivelser	(Landbrug\$ or skovbrug\$ or fiskeri\$ or gartneri\$ or famn\$ or forest\$ or fishing or gardener or landwirt\$ or forstwirt\$ or fischer\$ or gärtner\$) near2 (institution\$ or school\$ or \$skole\$ or \$schul\$)	0	0

PRÆCISERINGSMULIGHEDER I ET PARADOKS-INDEKSERET OG I ET TRADITIONELT KATALOG

Dette appendiks indeholder resultaterne af 25 sammenligninger mellem det PARADOKS-indekserede katalogs præciseringsmuligheder og Danmarks Pædagogiske Biblioteks online katalogs præciseringsmuligheder efter at der er foretaget klassifikationssøgninger, henholdsvis søgninger på klassernes navne.

Alle sammenligninger er dubleret således, at der både sammenlignes med det fulde PARADOKS-indekserede katalog og den del, der kun indekserer bøger.

Tabellen angiver antallet af fund i Danmarks Pædagogiske Biblioteks katalog (DPB) og de to PARADOKS-indekserede kataloger samt tidsskriftkataloget ERIC/CIJE.

Dernæst anføres antallet af forskellige ord fra søgningen i DPB's base, som samtidig optræder i nyere abstracts om emnet, taget fra ERIC.

Alle efterfølgende tal er beregnet i forhold til antallet af forskellige relevante ord i DPB-søgningerne, således at 30% betyder en ordmængde, der er 30% større end den oprindelige.

Først tælles hvor mange forskellige relevante ord søgesættet fra det PARADOKS-indekserede katalog mangler i forhold til søgesættet fra søgning i DPB's katalog. Dette opføres i kolonnen, kaldt tab.

Dernæst tælles hvor mange forskellige relevante ord søgesættet fra det PARADOKS-indekserede katalog har flere end søgesættet fra søgning DPB's katalog.

Dette opføres i kolonne, kaldet gevinst.

Endelig subtraheres de to størrelser og resultatet fremkommer. Er den positiv er det en reel gevinst, mens et negativt resultat betyder et reelt tab.

Emne	Antal poster			Ord- længde	Antal rele- vante ord DPB	Tab		Gevinst		Resultat	
	DPB		PARADOKS			PARADOKS		PARADOKS		PARADOKS	
		Alle	Bøger			Alle	Bøger	Alle	Bøger	Alle	Bøger
Biologi. Me- todik	7	14	11	41	35	31%	31%	306%	183%	274%	151%
					35	31%	31%	320%	189%	289%	157%
					37	30%	30%	305%	173%	276%	143%
					39	23%	28%	290%	154%	267%	126%
Datalogi. Metodik	39	10	6	35	86	44%	49%	64%	48%	20%	-1%
					87	44%	48%	64%	49%	21%	1%
					89	45%	49%	70%	52%	25%	2%
					105	41%	47%	65%	46%	24%	-1%
Matematik- vanskelig- heder	14	8	8	32	40	43%	43%	165%	165%	123%	123%
					37	38%	38%	189%	189%	151%	151%
					37	38%	38%	203%	203%	165%	165%
					37	35%	35%	205%	205%	170%	170%

Emne	Antal poster			Ord- længde	Antal rele- vante ord DPB	Tab		Gevinst		Resultat		
	DPB	PARADOKS				PARADOKS		PARADOKS		PARADOKS		
		Alle	Bøger			Alle	Bøger	Alle	Bøger	Alle	Bøger	
Elektronik og elektroteknik. Metodik	1	6	4	16	Fuld	1	0%	0%	3100%	2800%	3100%	2800%
					9	2	50%	50%	1700%	1600%	1650%	1550%
					7	2	0%	0%	1800%	1650%	1800%	1650%
					5	2	0%	0%	2450%	2250%	2450%	2250%
	5	3	2	20	Fuld	16	25%	31%	294%	213%	269%	181%
Energi. Me- todik					9	17	24%	29%	259%	188%	235%	159%
					7	17	18%	29%	276%	200%	259%	171%
					5	21	33%	43%	271%	205%	238%	162%
	14	37	26	44	Fuld	40	25%	30%	375%	330%	350%	300%
Fysik. Meto- dik					9	42	24%	29%	376%	329%	352%	300%
					7	45	18%	22%	356%	311%	338%	289%
					5	53	21%	25%	309%	268%	289%	243%

Emne	Antal poster				Ord- længde	Antal rele- vante ord DPB	Tab		Gevinst		Resultat	
	DPB	PARADOKS		ERIC			PARADOKS		PARADOKS		PARADOKS	
		Alle	Bøger				Alle	Bøger	Alle	Bøger		
Geografi. Metodik	6	20	9	77	Fuld	23	22%	26%	722%	422%	700%	396%
					9	25	28%	32%	696%	400%	668%	368%
					7	27	26%	33%	670%	400%	644%	367%
					5	29	21%	28%	641%	393%	621%	366%
Geologi. Metodik	1	8	2	51	Fuld	5	20%	60%	1240%	920%	1220%	860%
					9	5	20%	60%	1300%	960%	1280%	900%
					7	5	20%	60%	1320%	980%	1300%	920%
					5	5	20%	40%	1440%	1020%	1420%	980%
Husholdning. Metodik	4	1	1	27	Fuld	8	27%	27%	0%	0%	-27%	-27%
					9	9	35%	35%	0%	0%	-35%	-35%
					7	10	32%	32%	0%	0%	-32%	-32%
					5	15	41%	41%	0%	0%	-41%	-41%

Emne	Antal poster			Ord- længde	Antal rele- vante ord DPB	Tab		Gevinst		Resultat	
	DPB		PARADOKS			PARADOKS		PARADOKS		PARADOKS	
		Alle	Bøger			Alle	Bøger	Alle	Bøger	Alle	Bøger
Dataskin- styret under- visning	149	71	52	199	342	24%	25%	111%	102%	87%	76%
					345	21%	23%	108%	98%	86%	75%
					335	20%	23%	103%	96%	83%	73%
					326	20%	21%	93%	89%	73%	67%
Kemi. Meto- dik	6	15	9	53	35	31%	31%	317%	277%	286%	246%
					35	26%	29%	329%	286%	303%	257%
					37	24%	27%	308%	270%	284%	243%
					39	26%	31%	331%	297%	305%	267%
Landbrug- sundervisning	3	3	1	51	10	40%	70%	420%	50%	380%	-20%
					9	44%	56%	467%	67%	422%	11%
					9	50%	63%	550%	88%	500%	25%
					10	67%	80%	653%	93%	587%	13%

Emne	Antal poster				Ord- længde	Antal rele- vante ord DPB	Tab		Gevinst		Resultat	
	DPB	PARADOKS		ERIC			PARADOKS		PARADOKS		PARADOKS	
		Alle	Bøger				Alle	Bøger	Alle	Bøger		
Madlavning. Metodik	2	4	3	16	Fuld	6	50%	50%	300%	283%	250%	233%
					9	7	57%	57%	314%	300%	257%	243%
					7	7	57%	57%	343%	329%	286%	271%
					5	9	56%	56%	356%	333%	300%	278%
Matematik. Metodik	113	143	87	199	Fuld	264	18%	17%	282%	255%	264%	238%
					9	270	9%	10%	271%	243%	262%	233%
					7	271	8%	9%	244%	224%	236%	215%
					5	271	4%	6%	210%	193%	206%	187%
Matematik og samfund. Metodik	3	19	11	16	Fuld	4	25%	25%	2550%	2250%	2525%	2225%
					9	4	25%	25%	2650%	2300%	2625%	2275%
					7	4	25%	25%	2725%	2350%	2700%	2325%
					5	8	25%	25%	1475%	1338%	1450%	1313%

Emne	Antal poster			Ord- længde	Antal rele- vante ord DPB	Tab		Gevinst		Resultat	
	PARADOKS		PARADOKS			PARADOKS		PARADOKS			
	DPB	Bøger	Alle			Bøger	Alle	Bøger	Alle	Bøger	
Natur- videnskab og samfund	6	3	1	22	18	61%	78%	100%	22%	39%	-56%
					17	53%	71%	118%	24%	65%	-47%
					18	56%	72%	100%	22%	44%	-50%
					20	55%	70%	85%	20%	30%	-50%
Ernærings- lære. Metodik	14	2	2	50	61	10%	10%	10%	10%	0%	0%
					62	12%	12%	10%	10%	-2%	-2%
					68	12%	12%	11%	11%	-2%	-2%
					75	15%	15%	13%	13%	-3%	-3%
Teknologi og samfund. Metodik	3	26	18	13	10	10%	10%	870%	760%	860%	750%
					10	0%	0%	970%	830%	970%	830%
					11	0%	0%	882%	764%	882%	764%
					12	0%	0%	833%	750%	833%	750%

Emne	Antal poster			Ord- længde	Antal rele- vante ord DPB	Tab		Gevinst		Resultat	
	DPB		PARADOKS			PARADOKS		PARADOKS		PARADOKS	
		Alle	Bøger			Alle	Bøger	Alle	Bøger	Alle	Bøger
Færdselslære. Metodik.	1	3	3	9	Fuld	9	44%	200%	200%	156%	156%
					9	9	44%	211%	211%	167%	167%
					7	9	44%	256%	256%	211%	211%
					5	9	44%	300%	300%	256%	256%
Økologi. Metodik	11	26	11	174	Fuld	60	38%	572%	440%	533%	402%
					9	58	36%	614%	467%	578%	428%
					7	63	35%	570%	440%	535%	400%
					5	70	26%	501%	383%	476%	350%
Naturkund- skab. Metodik	89	14	8	199	Fuld	281	56%	60%	47%	4%	-14%
					9	286	55%	60%	47%	6%	-12%
					7	293	54%	55%	42%	1%	-16%
					5	295	51%	53%	41%	2%	-13%

Emne	Antal poster			Ord- længde	Antal rele- vante ord DPB	Tab		Gevinst		Resultat		
	DPB	PARADOKS				PARADOKS		PARADOKS		PARADOKS		
		Alle	Bøger			Alle	Bøger	Alle	Bøger			
Auditive hjælpemidler. Metodik.	9	5	1	37	Fuld	22	23%	82%	300%	23%	277%	-59%
					9	22	23%	82%	323%	23%	300%	-59%
					7	24	25%	79%	325%	21%	300%	-58%
					5	24	29%	79%	375%	25%	346%	-54%
Mediepæda- gogik	19	25	10	47	Fuld	35	20%	34%	397%	214%	377%	180%
					9	37	19%	41%	373%	203%	354%	162%
					7	43	16%	40%	349%	195%	333%	156%
					5	55	9%	31%	295%	184%	285%	153%
Teksttilvirk- ning, håndar- bejde	1	3	3	5	Fuld	4	0%	0%	250%	250%	250%	250%
					9	5	0%	0%	220%	220%	220%	220%
					7	6	0%	0%	167%	167%	167%	167%
					5	8	0%	0%	138%	138%	138%	138%

Emne	Antal poster				Ord- længde	Antal rele- vante ord DPB	Tab		Gevinst		Resultat	
	DPB	PARADOKS		ERIC			PARADOKS		PARADOKS		PARADOKS	
		Alle	Bøger				Alle	Bøger	Alle	Bøger		
Byggeteknik, arkitektur	1	3	2	7		4	50%	50%	125%	100%	75%	50%
						4	50%	50%	125%	100%	75%	50%
						5	60%	60%	120%	100%	60%	40%
					5	71%	71%	114%	100%	43%	29%	

REFERENCER

- Arley, Niels & K.Rander Buch: Sandsynlighedsregning og statistik. København: Gjellerup, 1961.
- Atherton, Pauline: Books are for use. Final report of the Subject Access Project to the Council on Library Ressources. Syracuse, 1977.
- Badiou, Alain: Le concepte de Modèle. Paris, 1969.
- Bates, Marcia J.: System meets user: Problems in matching subject search terms. *Information Processing and Management*, 13(6), 1977: 367-375.
- Bates, Marcia J.: Subject Access in Online Catalogs: A Design Model. *Journal of the American Society for Information Science*, 37(6), 1986: 357-376.
- Bates, Marcia J.: How to use information search tactics online. *Online*, 1987(may): 47-54.
- Bates, Marcia J.: The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5), 1989: 407-424.
- Bayer, Alan E. & Gerald Jahoda: Effects of Online Bibliographic Searching on Scientists' Informations Style. *Online Review*, 5, 1981: 323-33.
- Bernsen, Niels Ole & Ib Ulbæk: Naturlig og kunstig intelligens. Introduktion til kognitionsforskningen. København: Nyt Nordisk Forlag, 1993.
- Bhaskar, Roy: A realist theory of science. Leeds: Leeds Books, 1975.
- Bredsdorff, Nils & Claus Poulsen: Stikprøveanalyser af antallet af samleværker og antallet af artikler i samleværker indenfor det samfundsvidenskabelige og det naturvidenskabelige område på Roskilde Universitetsbibliotek, 1993. Memo.
- Brunak, Søren & Benny Lautrup: Neurale netværk. Computere med intuition. København: Nysyn/Munksgaard, 1988.
- Byrne, Alex & Mary Micco: Improving OPAC Subject Access: The ADFA Experiment. *College & Research Libraries*, 49(5), 1988: 434 ff..

Byrne, Alex: Life wasn't meant to be whimsical: Painless subject augmentation. *Australasian College Libraries*, 4(2), 1986: 83-90.

CAS, American Chemical Society: Statistical Summary 1907-1992. Ohio: CAS, 1993.

Cavallin, Mats: Kapitelindex. Att utöka katalogposter med kapitelrubriker, ämnesindex och bibliografier. Slutrapport från ett BIBSAM-projekt. Göteborg Universitetsbibliotek, Göteborg, 1991. (BIBSAM-rapport nr.4.)

Chartron, Ghislaine; Dalbin, Sylvie; Monteil, MarieGaeel; Verillon, Monique: Manual and automatic indexing: beyond the differences. /Indexation manuelle et indexation automatique: dépasser les oppositions. *Documentaliste*, 26(4-5), 1989: 181-187.

Cleverdon, Cyril, Jack Mills & Michael Keen: Factors determining the performance of indexing systems. Vol.I, Design; part.I, Text. Aslib-Cranfield Research Project. Cranfield Institute of Technology, Cranfield, England, 1966.

Cochrane, Pauline Atherton: Redesign of Catalogs and Indexes for Improved Online Subject Access. Selected Papers of Pauline A. Cochrane. Phoenix, Az, 1985.

Cooper, William S.: Is Interindexer Consistency a Hobgoblin? *American Documentation*, 1969(july): 268-278.

Cooper, W.S.: Getting beyond Boole. *Information processing and management*, 23, 1988: 243-248.

Davison, P.S. et al.: International Bibliographic Review on Costs and Modelling in Information Retrieval: A Report on Completion of Data Collection, Assessment of Efficiency of Searching, and on Means of Indexing This Information. London: British Library, 1988.

Det kongelige bibliotek i forandring. Moderniseringsprojektet 1986-88. Det Kongelige Bibliotek. København, 1989. (Modernisering i Historiske Rammer No. 3.)

Diodato, Virgil: Tables of contents and book indexes: how well do they match reader's descriptions of books? *Library Resources & Technical Services*, 30(4), 1986: 402-412.

Fried, Charles and Prevel, James J.: Effects of Indexing Aids on Indexing Performance. Bethesda, Md., General Electric Company, Oct. 1966. Refereret fra Leonard, 1977.

Fugmann, Robert: The Navigational Index. *International Classification*, 18(2), 1991: 122-132.

Gabel, Dorothy L., Martin H. Kagan & Robert D. Sherwood: A Summary of Research in Science Education - 1978. *Science Education*, 64(4), 1980: 483.

Gale directory of databases. 1992-.

Hawkins, Brian L.: Planning for the National Electronic Library. *Educom Review*, (may/june), 1994:19-29.

Henzler, Rolf G.: Free or Controlled Vocabularies. Some statistical user-oriented evaluations of biomedical informationsystems. *International Classification*, 5(1), 1978: 21-32.

Herr, J.J.: Use of Data-base Access for Interindexer Communication and for Indexer Training. *Proc. ASIS*, (33'th ann. meeting), 1970.

Hjørland, Birger: Emnerepræsentation og informationssøgning : Bidrag til en teori på kundskabsteoretisk grundlag. Göteborg, 1993. (Skrifter från VALFRID 4)

Hoffman, Herbert H. and Jeruel L. Magner: Future Outlook: Better Retrieval through Analytic Catalogs. *The Journal of Academic Librarianship*, 11(2), 1985: 151-153.

Houston, James E. ed.: Thesaurus of ERIC descriptors; 12. ed. Phoenix, Ariz. : Oryx Press, 1990.

Hyman, Richard J.: Analytical access: history, resources, needs. Flushing, New York, Queens College of the City University of New York, 1978. (Queens College Studies in Librarianship.)

Høy Nielsen, Inger, Ingar Lomheim & Irma Pasanen-Tuomainen: Monitoring Online Catalogues (OPACs) in the Nordic Technological University Libraries. *NORDINFO--publikation*, 23, 1992.

Iivonen, Mirja: Interindexer consistency and the indexing environment. *International Forum on Information and Documentation*, 15(2), 1990:16-21.

ISO standard 2788.

Knightly, John J.: Traditional Information Gathering vs. the Computer: Surprising Findings. *Tennessee Librarian*, 31(3), 1979: 32-38.

Lancaster, Wilf: Natural languages versus controlled languages; a thirty year review of the literature of information science. In: Oppenheim et al. (Eds.): Perspectives in information management vol. 1. London, Butterworth. 1989.
Reprinted in: Lancaster, Wilf: Indexing and abstracting in theory and practice. 1991: 193-218.

Lancaster, F.W.: Indexing and Abstracting in Theory and Practice. London: The Library Association, 1991

Lancaster, F.W., Tschera Harness Connell, Nancy Bishop & Sherry McCowan: Identifying barriers to effective subject access in library catalogs. *Library Resources and Technical Services*, 35(4), 1991: 377-391.

Larson, Ray. R.: Workload Characteristics and Computer System Utilization in Online Library Catalogs. Doctorial Dissertation, University of California at Berkeley, 1986. (University Microfilms No. 8624828.)
Citeret efter Larson 1991.

Larson, Ray. R.: The decline of subject searching: long-term trends and patterns of index use in an online catalog. *Journal of American Society for Information Science*, 42(3), 1991: 197-215.

Leonard, Lawrence E.: Inter-Indexer Consistency and Retrieval Effectiveness: Measurement of Relationships. Ph.D. dissertation prepared for the University of Illinois Graduate School of Library Science, 1975.
Citeret after Leonard 1977.

Leonard, Lawrence E.: Inter-indexer consistency studies, 1954-1975: A review of the literature and summary of study results. University of Illinois. Graduate School of Library Science. 1977. (Occasional papers no. 131.)

Lewis, David W.: Research on the Use of Online Catalogs and Its Implications for Library Practice. *Journal of Academic Librarianship*, 13(3), 1987: 152-157.

MacMillan, J.T. & I.D.Welt: Study of indexing procedures in a limited area of the medical sciences. *American Documentation*, 12, 1961: 27-31.

Maizell, R.E.: Value of Titles for Indexing Purposes. *Revue de la Documentation*, 27, 1960: 126-127.

Mandel, Carol A.: Enriching the Library Catalog Record for Subject Access. *Library Resources and Technical Services*, 29(1), 1985: 5-15.

Markey, Karen: Analytical review of catalog use studies. Columbus, OH: OCLC, 1980. (Technical Report Number OCLC/OPR/RR-80/2.)

Markey, Karen: Online Catalog Use: Results of Surveys and Focus Group Interviews in Several Libraries. Vol 2. of the Final Report to the Council on Library Resources. Dublin, OH: Office of Research, OCLC Online Computer Library Center, 1983: 141 ff..

Markey, Karen: Subject Searching in Library Catalogs: Before and After the Introduction of Online Catalogs. Dublin, OH: OCLC, Inc., 1984.

Markey, Karen: Subject-Searching Experiences and needs of online catalog users: Implications for Library Classification. *Library Resources & Technical Services*, 29(1), 1985: 34-51.

Markey, Karen: Class number searching in an experimental online catalog. *International Classification*, 13(3), 1986: 142-150.

Markey, Karen: Interindexer consistency tests: a literature review and report of a test of consistency in indexing visual materials. *Library and Information Science Research*, 6(2), 1984: 155-177.

Markey, Karen: Integrating machine-readable LCSH into online catalogs. *Information Technology and Libraries*, 7, 1988: 299-312.

Markey, Karen & Karen Calhoun: Unique Words Contributed by MARC Records with Summary and/or Content Notes. *Proc. ASIS*, 24(50'th ann. meeting), 1987: 153-162.

Matthews, J.R. & Gary S.Lawrence: Further analysis of the CLR Online Catalog Project. *Information Technology and Libraries*, 3, 1984: 354-371.

Metzger, H: Newton, Stahl, Boerhave et la doctrine chimique. Paris, 1930.

- Posey, Edwin & Charlotte Erdmann: An Online UNIX-Based Engineering Library Catalog: Purdue University Engineering Library. *Science & Technology Libraries*, 6(4), 1986: 32 ff..
- Poulsen, Claus: Subject Access to New Subjects, Specific Paradigms and Surveys: PARADOKS-registration. *LIBRI*, 40(3), 1990: 179-202.
- Poulsen, Claus: Online kataloger med indholdsfortegnelser af samleværker I: Kvalitetsforbedring for brugeren ved omlægning af ressourceforbruget. *Biblioteksarbejde*, 13(36), 1992: 31-42.
- Poulsen, Claus: Papir, CD-rom og online. Løbende bibliografier inden for matematik, datalogi, fysik, energi og naturvidenskab generelt. Skriftserie fra Roskilde Universitetsbibliotek, 19, 1993.
- Poulsen, Claus: Højspænding, elektromagnetiske felter og sundhed. Skriftserie fra Roskilde Universitetsbibliotek, 21, 1993.
- Poulsen, Claus & Anni Lenschau Teglers: Online kataloger med indholdsfortegnelser af samleværker II: Undersøgelse af Danmarks Pædagogiske Bibliotek. *Biblioteksarbejde*, 13(36), 1992: 43-47.
- Pöttsch, Marianne: Indexiertiefe und Uebereinstimmung der Ergebnisse bei Verwendung modifizierter Richtlinien. *Informatik*, 27(2), 1980: 29-30.
- Reich, Phyllis & Erik J.Biever: Indexing consistency: the input/output function of thesauri. *College & Research Libraries*, 52(4), 1991: 336-342.
- Rodgers, Dorothy: A Study of Inter-Indexer Consistency. Washington D.C.: General Electric Co., 1961.
Refereret fra Leonard, 1977
- Rolling, R.: Indexing consistency, Quality and Efficiency. *Information Processing & Management*, 17(2), 1981: 69-76.
- Rumelhart, David E. & James L. McClelland and the PDP Research Group: Parallel distributed processing. Explorations in the Microstructure of Cognition. Vol.1: Foundations. 9. pr. MIT Press. Cambridge Mass., 1989.
- Salton, Gerald: Automatic Text Processing. The transformation, analysis and retrieval of information by computer. Addison-Wesley, Reading, Mass., 1989
- Saracevic, Tefko Measuring the Degree of Agreement between Searchers. *Proc. ASIS*, 21(47'th ann. meeting), 1984: 227-230.
- Slamecka, V. & J.J.Jacoby: Consistency of human indexing. In: Vladimir Slamecka (ed.): The Coming Age of Information Technology. Documentation Inc., Bethesda, MD, 1965: 32-56.
- Söderquist, Thomas (1994): Who Will Sort Out the Hundred or More Paul Ehrlichs? - Remarks on the Historiography of Contemporary and Recent Science. Background Paper for Working Conference "The Historiography of Contemporary History of Science, Technology, and Medicine", Göteborg Univeristy, 16-17 Sepetember 1994. (To be published)

Tarr, Daniel & Harold Borko: Factors influencing inter-indexer consistency. *Proc. ASIS*, 11(37'th ann. meeting), 1974: 50-55.

Tell, Björn V.: Document representation and indexer consistency. - A study of indexing from titles, abstracts and full text using UDC and keywords. *Proc. ASIS*, (32'th ann. meeting), 1969: 285-292.

Walker, Stephen & Richard M Jones: Improving subject retrieval in online catalogues: 1. Stemming, automatic spelling correction and cross-reference tables. *British Library Research Papers*, 24, 1987.

Wiberley, S.E. & Daugherty, R.A.: Users' persistence in scanning lists of references. *College and Research Libraries*, 49, 1988: 149-156.

Wormell, Irene: Subject Access Project - SAP: Improved Subject Retrieval for Monographic Publications. Lund: Lund University, 1985.

Zunde, Pranas & Margeret E. Dexter: Factors affecting indexing performance. *Proc. ASIS*, (32'th meeting), 1969: 313-32.

STIKORDSREGISTER

Nedenstående består dels af ord fra mellemrubrikker, anvendt i bogen, dels væsentlige definitioner og henvisninger i teksten. Endelig er anvendte referencers første forfatter anført i registret, som hjælp til at lokalisere referenceanvendelsen.

- | | | | |
|-----------------------------------|---------------------------|---------------------------------------|--------|
| Affinitet | 61 | Emnebegreb | |
| Aksiomer | 59 | fagvidenskabeligt | 55, 56 |
| Analytiske kataloger | 89, 92 | informations- og dokumentations- | |
| omfang | 95 | videnskabeligt | 53, 56 |
| resursekrav | 94 | Emnebeskrivelsernes kvalitet | |
| ressurser | 103 | forbedringer | 22 |
| Arley | 153, 191 | Emnedatasystemer | 63 |
| Atherton | 90-92, 103, 191, 192 | Emneindekseringens længde . . . | 31 |
| Badiou | 70, 191 | Emnesøgning ved "browsing", | |
| Bates | 4, 9, 16, 18, 19, 84, 191 | græsning eller bladrning | 6 |
| Bayer | 10, 191 | Emnesøgning ved gennemgang af | |
| Bernsen | 77, 191 | kernetidsskrifter | 6 |
| Bhaskar | 51, 52, 191 | Emnesøgning via direkte henvende- | |
| Bibliotekskatalogers anvendelse . | 10 | lse til emnespecialister | 7 |
| udvikling | 11 | Emnesøgning via forfattersøgning | 7 |
| Bredsdorff | 96, 100, 101, 191 | Emnesøgning via oversigtslittera- | |
| Brugernes ønsker | 25 | tur | 8 |
| problemer | 26 | Emnesøgning via personlige kon- | |
| undersøgelser | 25 | takter | 7 |
| Brunak | 77, 191 | Emnesøgning via referencer | 7 |
| Byrne | 27, 91, 191, 192 | Emnesøgning via tidsskriftsbibliogra- | |
| Båndbredde for læsning | 76 | fier | 5 |
| Båndbredde for tolkning | 76 | Emnesøgningens metoder | 3 |
| CAS | vi, 192 | Emnesøgningernes kvalitet | 19 |
| Cavallin | 94, 103, 104, 192 | forbedringer | 21 |
| Chartron | 32, 192 | Emnesøgningsmuligheder | |
| Cleverdon | 30, 192 | forbedring af emnebeskrivelsen | 18 |
| Cochrane | 90-92, 103, 192 | forbedring af søgeprocessen . . | 18 |
| Cooper | 18, 44-46, 192 | Emnesøgningsmuligheder, forbe- | |
| Dalton | 61 | dring | 16 |
| Davison | 8, 192 | Emnesøgning i bibliotekskataloger | 4 |
| Diodato | 19, 192 | | |

- Euklid 59
 Evaluering 115
 emneord 119, 121
 klassifikationskoder ... 119, 123
 præciseringsmuligheder 119
 reproducérbarhed 120
 struktur 117
 Formelt system 59
 Fri tekst 30
 Fried 37, 192
 Fugmann . . . iii, 30, 32, 34, 35, 192
 Fuldstændighed 72
 Fysiske blandinger 61
 Gabel 29, 192
 Gale 6, 192
 Generelle begreber 32
 Geometri 59
 Henzler 35, 42, 121
 Hjørland iii, 54
 Iivonen 39, 40, 42, 193
 Indekseringmetode
 indholdsorienteret 53
 Indekseringskonsistens 35
 genfinding 43
 mål 41
 Indekseringsmetode
 behovsorienteret 53
 Informationsindhold på system-
 niveau 74
 ISO 64, 193
 Kemi 61
 Kemiske forbindelser 61
 Klassifikationssystemer 63
 Knightly 10, 193
 Konservative begreber 33
 Konsistens 69
 Konsistensmål 41
 Kontrollerede sprog 29, 31
 Kontrolleret sprog 31
 Lancaster . . . 19, 35, 53, 147-149,
 148, 149, 193
 Larson 11-14, 13, 16, 17, 193
 Leonard 36-39, 42-44, 46,
 192, 194, 195
 Lukkede systemer 52
 MacMillan 39, 194
 Maizell 121, 194
 Marginal litteratur 33
 Markey iii, 10, 11, 13, 16, 18,
 22-24, 26, 36, 38, 40, 43, 44, 194
 Matthews 13, 194
 Metzger 61, 194
 Model 59, 72
 Modelbegreb 65
 Mætning 70
 Naturlig tekst 29-31, 89
 Nyskabende litteratur 33
 Oversigtslitteratur 8
 Paradoks-kataloget, Danmarks
 Pædagogiske Bibliotek
 dokumenttyper 110
 publikationsår 109
 sproglig fordeling 112
 udvalgskriterier 111
 Paradoksindeksering
 udformning 106
 Poulsen ii, iii, 5, 9, 33,
 96-101, 191, 195
 Præciseringsmuligheder 125
 Pötzsch 195
 Realistisk ontologi 51
 Redundans 32, 76
 Reich 38, 39, 195
 Relevante ord 126
 Rodgers 37, 195
 Rolling 42, 43, 45, 195
 Rumelhart 77, 195
 Salton 18, 195
 Samleværker 92
 Samleværker, analyse
 Danmarks Pædagogiske Biblio-
 tek 96
 Roskilde Universitetsbibliotek 100
 Santa Ana College Library . . . 96
 SAP-indeksering 90
 Saracevic 43, 195
 Semantiske grundbegreber 70
 Semantiske normer 72
 Slamecka 38, 195
 Subject Access Project 90
 Syntaktiske grundbegreber 66
 Syntaktiske normer 69
 Syntetiserende begreber 32
 Söderquist 9, 195
 Tarr 38, 195
 Teoretisk system 59
 Thesauri 64
 Tolkning 59
 Transcendental epistemologi . . . 51
 Transcendental realisme 52
 Tvetydige emnebeskrivelser . . . 34
 Utidssvarende emnebeskrivelser 33
 Videnskabsteoretiske grundbegre-
 ber 50
 Walker 18, 196
 Wiberley 16, 196
 Zunde 36, 37, 39, 40, 196